

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 21:14:07 ; Search time 1659.36 seconds
(without alignments)
18464.115 Million cell updates/sec

Title: US-09-664-186-4

Perfect score: 1026

Sequence: 1 GGAAGAGCAAGAAACCTT.....CCCTCAAGAGGCTCTCTGA 1026

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pbg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	5.8	683	14	CA235176
C 2	54.6	5.3	925	29	CNS0091P
3	53.2	5.2	655	29	CG352167
C 4	53.2	5.2	679	29	CG609109

C	5	53.2	5.2	773	29	CG661222	OGLAU18TV
	6	53.2	5.2	807	28	BZ529007	OGAIR91TC
	7	53.2	5.2	815	29	CG034209	PUGY50TD
	8	53.2	5.2	897	29	CG661216	OGLAU18TH
	9	51.8	5.0	925	29	CNS0091P	AL053013
	10	51.2	5.0	649	14	CA230834	SCUFFL3C0
	11	51.2	5.0	649	14	CB870888	HC15J04Y
C	12	50.6	4.9	594	14	CB874255	BI960118
	13	50.6	4.9	637	12	BQ762267	HVSME002
	14	50.6	4.9	664	13	BQ762267	EBT001 SQ
	15	50.6	4.9	973	14	CA275829	CA275829
	16	50.4	4.9	855	29	CG041883	SCCSD109
	17	50.4	4.9	890	29	CG731258	PUEP24TD
	18	50.4	4.9	937	29	CC29649	CG731258
	19	50.2	4.9	509	12	BJ549094	OGLAJ02TV
	20	49.8	4.9	608	14	CA179624	BJ549094
	21	49.8	4.9	659	14	CA194749	CA179624
	22	49.8	4.9	694	14	CA196087	SCVPS106
	23	49.4	4.8	728	14	CG043714	CA194749
C	24	49.4	4.8	775	29	CG207191	SCRF8106
	25	49.4	4.8	793	29	CG449321	CA196087
	26	49.4	4.8	837	29	CG701337	SCSEAD108
C	27	49.4	4.8	1201	13	BX381961	CA194749
	28	49.2	4.8	502	12	BM335757	SCVPS106
	29	49.2	4.8	728	14	CF638161	CA194749
	30	49.2	4.8	759	10	BP259495	SCRF8106
	31	49.4	4.8	745	29	CG633883	CA196087
C	32	49.4	4.8	869	29	CG284200	CA194749
	33	49.4	4.8	909	29	CG228851	SCRF8106
C	34	49.4	4.8	959	14	CA208502	CA196087
	35	48.6	4.7	559	14	CA210811	SCVPS106
	36	48.6	4.7	564	14	CA210811	SCVPS106
	37	48.6	4.7	605	14	CA181598	SCVPS106
	38	48.6	4.7	620	14	CA182549	SCVPS106
	39	48.6	4.7	647	14	CA216848	SCVPS106
	40	48.6	4.7	730	14	CA195888	SCVPS106
C	41	48.6	4.7	902	28	BZ579152	CA195888
	42	48.4	4.7	524	9	AV436476	BZ579152
C	43	48.4	4.7	807	29	CG731270	AV436476
	44	48.4	4.7	910	29	CG323065	CG731270
C	45	48.4	4.7	1066	14	CA254564	CG323065

ALIGNMENTS

RESULT 1
CA235176
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CA235176
SCCCFL4093A04.g FL4 Saccharum officinarum cDNA clone
5', mRNA sequence.
CA235176
CA235176.1 GI:35303456
EST.
Saccharum officinarum
Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 683)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
The libraries that made SUCEST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccetec.fcap.unesp.br

Plate: 093 row: A column: 04
 Seq primer: T7 Promoter Primer.
 Location/Qualifiers
 1. .683
 /organism="Saccharum officinarum"
 /mol_type="mRNA"
 /db_xref="taxon:4547"
 /clone="SCCFLA093A04"
 /lab_host="DH10B"
 /clone_lib="F14"
 /note="Organ: Developed inflorescence and rachis (20cm-long); Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from (Developed inflorescence and rachis (20cm-long)). cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at
 http://sucest.lad.ic.unicamp.br/public"

ORIGIN
 Query Match 5.8%; Score 60; DB 14; Length 683;
 Best Local Similarity 46.7%; Pred. No. 0.088;
 Matches 189; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

357 CAACCGGAGCTCTGGAGGAGTGGGAGCGTCTCTTCATGTCCTCCCTCGAGATGT 416
 Db
 254 CATCTCTCTCAGCGCTGACCGTGGCGCGCGCGGTGGCGGTGCGCGTCAATCGCGC 313
 Qy
 417 GGCCTCAACCTGGGGGTACCCCGGAGACCGTCCACCGCTGGAAGAAGTCTTGAAGA 476
 Db
 314 GCCGACCCCGCGCGCGCGTGTACACCGTCCACATGTCGGCACCGTGTCTCTC 373
 Qy
 477 AAGGGCTGTGGTCCACGACGCTCTTACMAACCGTCAACGGGAGCGCGGGCCAT 536
 Db
 374 CATCTGCAAGGCGCGCGCGCGTGTCTCGCTTCTCCCGACCGCGGAGTCTCTCGA 433
 Qy
 537 CGCACACCTTTGGCGCGTCCGGTGGAGCCAGGAAAGCAGGCTCACCTCGACACTA 596
 Db
 434 CGGCTCAGTCTAGTCTGCGGAGGAGCGCGCGTCACTCTCGCATGTTGGCGG 493
 Qy
 597 CATCTACCTCGAGGAACCTCGCTTAGACATGCGCAACCGCGTCTCTCTCAACTG 656
 Db
 494 GCTCGCGTGCCTATCTTCTGCTCAGTGGGTGCGGCTCGCGCTCGCTTGTGTCAG 553
 Qy
 657 GGTCAAGGCTACAGGACACAGGATCGCCACCCCTGGACGTGCTGCTCTGGGC 716
 Db
 554 GTACTAGTCTAGTGTACAGGAGTGGCGGACACCCGATGCGCCACCGCCCAAG 613
 Qy
 717 TCAGGGGAAAGGTTGATCCCAACACCAAGACCGTGGCGTTGA 761
 Db
 614 TCGGGGGAAGGAACCGGCGCCACCTGGCCCTGGCCCTTCA 658

RESULT 2
 CDS0091P/c 925 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
 DEFINITION BAC19D16 of RPI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 AL053013
 AL053013.1 GI:4934461
 GSS.
 Drosophila melanogaster (fruit fly)
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 925)
 Genoscope.
 Direct Submission

FEATURES
 source
 1. 925
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BAC19D16"
 /clone_lib="RPI-98"
 /note="Tend : TET3"

COMMENT
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammos in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

Location/Qualifiers

ORIGIN
 Query Match 5.3%; Score 54.6; DB 29; Length 925;
 Best Local Similarity 15.8%; Pred. No. 1.2;
 Matches 63; Conservative 163; Mismatches 172; Indels 0; Gaps 0;

262 CCCCCCTTGGAGATGTATCAAGCGTCCAGAGCGGCTCAGAGCTTCTGGAACTC 321
 Db
 917 CSCSBSCSSSMSTSSNSBSCSSBSSTSSMSSSSSSSSSSSSSSSSSSSSSS 858
 Qy
 322 CTCAGGAGATTGCCCGCTCCACATCCCTCAGCAACCGGAGCTCTCGAGGAAGTG 381
 Db
 857 VKNASSSCGCGCMABCCMCWSSSSSCGASARGVKVRASGGAGRGGSGSASHS 758
 Qy
 382 GGGAGGTCTCTTCATGTCCTCCCTGGAGATGTTGGCCCTCAACCTGGGGTCACTCCG 441
 Db
 797 SSSACBSSSSCSACSWSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 738
 Qy
 442 CAGACCTCTCACCGCTCGAAGAGGCTCTTGAGAAAAAGGCGCTGTGGCCACCGACGTC 501
 Db
 737 SASSSSSSCSSSVSCSVASSMSCSBSBSBSASASSSSSSSSSSSSSSSSSSS 678
 Qy
 502 CTTACCAAAACCTCAACGGGAGCGCGGCCCATCGGCACCCCTTTGGCCGCTCGGCTG 561
 Db
 677 TSASMSAARSSS 618
 Qy
 562 AGGCCAGGAAAGCCAGGCTCACCTGGAGACTACATCTACCCCTGGAGAACCTCGCC 621
 Db
 517 GSVSASSGSSSVSSSGRSSGGGCGGSGSSSSSSSSSSSSSSSSSSSSSSSS 558
 Qy
 622 CTAGACATGCCAACCGCGTGTCTCTCTCAACTGGGT 659
 Db
 557 CSSAAAAAASCAVASCMMCGSKSGCTNNNNNTNTNT 520

RESULT 3
 CG352167 655 bp DNA linear GSS 26-AUG-2003
 CG352167 OG1BW12TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBma072A23,
 LOCUS genomic survey sequence.
 DEFINITION
 CG352167
 ACCESSION
 CG352167.1 GI:34269433
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Zea mays
 Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
AUTHORS

1 (bases 1 to 655)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE
JOURNAL
COMMENT

Consortium for Maize Genomics
Unpublished (2002)
Other_GSSs: OG1BW12TH
Contact: Cathy Whitelaw

TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..655

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBMA0727A23"

/clone_lib="ZM 0.7 1.5 KB"

/note="Vector: pBCSK; Site 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

ORIGIN

```
Query Match      5.2%; Score 53.2; DB 29; Length 655;
Best Local Similarity 46.1%; Pred. No. 2.1;
Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

Qy 321 CTTCCAGGAGATGCGCGCTCCACCATCCCTACGGCAACCGGAGCTCTGGAGGAGGT 380
Db 145 CTTCCAGGAGATGCGCGCTCCACCATCCCTACGGCAACCGGAGCTCTGGAGGAGGT 380

Qy 381 GGGGACGGTCTGCTTCATGGTCCCTCGAGATGTTGGCCCTCAACCTGGGGGTCAACCG 440
Db 205 GCTCAGGGCTGCTTCATGGTCCCTCGAGATGTTGGCCCTCAACCTGGGGGTCAACCG 440

Qy 441 CGAGACCGTTCAGGCTGGAAGAAGTCTTTAGAAAAGGCTTGGTGGCCACCGAGT 500
Db 265 CAGCTACCGGACACCGCTCGGGGCGGCTGCGGTACGGTTTCATCACCCATCGGCGG 324

Qy 501 CTTTACCAAAACCGTCAACGGGGAGCGCGGCGCATCGGCACCTTTGGGCGTCCGGCT 560
Db 325 CTTGCTCCCATCGAGGTGGCGGTCCAGCTCCGGCCCGCCCGCGCGCGCGCGC 384

Qy 561 GAGGCGAGGAAAGCGAGCTCACCTGGACGACTATCTATCCCTGGAGGAACCTCGC 620
Db 385 GCGGACGACAGGTACAGGCTGACCGTGGCGGACGCTATGACACGGGCTGCTGTTTCGC 444

Qy 621 CTTAGACATGGCAACGGGCTCTCTCTCAACTGGTTCAGGCTTACGAGGACACGG 680
Db 445 GGTGTTCTTGGCGTGGCCATGGTGGACGCAACGTGGTGGCGTCTTCTACCCCGTGA 504

Qy 681 AATCCGCCCGCCACCTCGAGCTGTGG 706
Db 505 GTCCGGCTCCACCATCAGCTGTGG 530
```

RESULT 4
CC609109/c

LOCUS

OGWDJ44TH ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0544H16,
genomic survey sequence.

ACCESSION

CC609109

VERSION

CC609109.1

GI:31970530

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

REFERENCE

AUTHORS

1 (bases 1 to 679)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

TITLE

JOURNAL

COMMENT

Consortium for Maize Genomics

Unpublished (2002)

Other_GSSs: OGWDJ44TV

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..679

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBMA0544H16"

/clone_lib="ZM 0.7 1.5 KB"

/note="Vector: pBCSK; Site 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

ORIGIN

```
Query Match      5.2%; Score 53.2; DB 29; Length 679;
Best Local Similarity 46.1%; Pred. No. 2.1;
Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;

Qy 321 CTTCCAGGAGATGCGCGCTCCACCATCCCTACGGCAACCGGAGCTCTGGAGGAGGT 380
Db 522 CTTCCAGGAGATGCGCGCTCCACCATCCCTACGGCAACCGGAGCTCTGGAGGAGGT 380

Qy 381 GGGGACGGTCTGCTTCATGGTCCCTCGAGATGTTGGCCCTCAACCTGGGGGTCAACCG 440
Db 562 GCTCAGGGCTGCTTCATGGTCCCTCGAGATGTTGGCCCTCAACCTGGGGGTCAACCG 440

Qy 441 CGAGACCGTTCAGGCTGGAAGAAGTCTTTAGAAAAGGCTTGGTGGCCACCGAGT 500
Db 502 CAGTACCGGACACCGCTCGGGGCGGCTGCGGTACGGTTTCATCACCCATCGGCGG 443

Qy 501 CTTTACCAAAACCGTCAACGGGGAGCGCGGCGCATCGGCACCTTTGGGCGTCCGGCT 560
Db 442 CTTGCTCCCATCGAGGTGGCGGTCCAGCTCCGGCCCGCCCGCGCGCGCGC 383

Qy 561 GAGGCGAGGAAAGCGAGCTCACCTGGACGACTATCTATCCCTGGAGGAACCTCGC 620
Db 382 GCGGACGACAGGTACAGGCTGACCGTGGCGGACGCTATGACACGGGCTGCTGTTTCGC 323

Qy 621 CTTAGACATGGCCAAACGGCGTGTCTCTCTCAACTGGGTCAAGGCTTACGAGGACCA 680
Db 322 GGTGTTCTTGGCGTGGCCATGGTGGACGCAACGTGGTGGCGTCTTCTACCCCGTGA 263

Qy 681 AATCCGCCCGCCACCTCGAGCTGTGG 706
Db 262 GTCCGGCTCCACCATCAGCTGTGG 237
```

RESULT 5
CC661222/c

LOCUS

OGLAU18TV ZM 0.7 1.5 KB Zea mays genomic clone ZMMBMA0312C11,
genomic survey sequence.

ACCESSION

CC661222

VERSION

CC661222.1

GI:32064977

KEYWORDS

GSS.

SOURCE

Zea mays

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T., Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T., Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGLAU18TH
Contact: Cathy Whitelaw
TIGR

7912 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

FEATURES

source

Location/Qualifiers

1..773

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBma0312C11"

/clone_lib="ZM 0.7-1.5 KB"

/note="vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

ORIGIN

Query Match 5.2%; Score 53.2; DB 29; Length 773;
Best Local Similarity 46.1%; Pred. No. 2.2; Indels 0; Gaps 0;
Matches 178; Conservative 0; Mismatches 208;

QY 321 CCTCCAGAGATTCCCGCTCCACCATCCCTACGGCAACCGGGAGCTCTGGAGGAAGT 380
DB 395 CTTCGAAATGCTGCGCGTCTTACCGGCTACGGTACTGTCAACCGCGGACCGCGC 336
QY 381 GGGACGGTCTGTATATGTCCTCCCTGGAGATGTTGGCCCTCAACTGGGGTACCCG 440
DB 335 GCTCAGGGGTGCTCATCGCGCTCTGCTGCTTCTGCTCTGCTCTGCTCTACCGA 276
QY 441 GCAGACCGTCCAGCTGGAGAGAGTCTTTGAGAAAAGGGCTGTGGTGGCCACGAGT 500
DB 275 CAGCTACCGGACACCGCCACGGGGGCTGCGGTACGGTTCATCACACCGGGCGG 216
QY 501 CTTTACCAACCGTCAACGGGAGCGCGGGCCATCGGACCCCTTTGGCGCTCGGCT 560
DB 215 CTTGCTCCCATCGAGGGTGGCGCTCCAGTCCGGCCCGCGCGCGCGCGCGC 156
QY 561 GAGCCAGGAAGACCGAGGCTCACCTGGAGAGTACATCTACCTGGAGAACCTCGC 620
DB 155 GCGGAGCAGCAGTACAGGCTACCGTGGCGGACGTATGCGAGGCTCTGTCTGCTGC 96
QY 621 CTTAGACATGCCAACCGGCTGTCTCTTCACTGGGTCAAGGCTCAACGAGCACCG 680
DB 95 GGTGTTCTTGGCGGTGCGCATGTGACCGACGTTGGTGGGTGTTCTTACCCCGTGA 36
QY 681 AATCCGCCCCCCTGGAGTGTGG 706
DB 35 GTCCCGTCCACCGAGCTGCTGG 10

RESULT 6

BZ529007

LOCUS

DEFINITION OGAIR917C ZM2_0.7-1.5_KB Zea mays genomic clone ZMMBma0070014, GSS 16-DEC-2002

genomic survey sequence.

ACCESSION BZ529007

VERSION BZ529007.1

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2002)

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Class: sheared ends.

FEATURES

source

Location/Qualifiers

1..807

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBma0070014"

/clone_lib="ZM2_0.7-1.5 KB"

/note="vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

ORIGIN

Query Match 5.2%; Score 53.2; DB 28; Length 807;
Best Local Similarity 46.1%; Pred. No. 2.2; Indels 0; Gaps 0;
Matches 178; Conservative 0; Mismatches 208;

QY 321 CCTCAGAGATTCCCGCTCCACCATCCCTACGGCAACCGGGAGCTCTGGAGGAAGT 380
DB 142 CTTTGAATGCTGCGCGTCTTACCGGCTACGGTACTGTCAACCGCGGACCGCGC 201
QY 381 GGGACGGTCTGTATATGTCCTCCCTGGAGATGTTGGCCCTCAACTGGGGTACCCG 440
DB 202 GCTCAGGGGTGCTCATCGCGCTCTGCTGCTTCTGCTCTGCTCTTACCGA 261
QY 441 GCAGACCGTCCAGCTGGAGAGAGTCTTTGAGAAAAGGGCTGTGGTGGCCACCGAGT 500
DB 262 CAGCTACCGGACACCGCCACGGGGGCTGCGGTACGGTTCATCACACCGGGCGG 321
QY 501 CTTTACCAACCGTCAACGGGAGCGCGGGCCATCGGACCCCTTTGGCGCTCGGCT 560
DB 322 CTTGCTCCCATCGAGGGTGGCGCTCCAGTCCGGCCCGCGCGCGCGCGC 381
QY 561 GAGCCAGGAAGACCGAGGCTCACCTGGAGAGTACATCTACCTGGAGAACCTCGC 620
DB 382 GCGGAGCAGCAGGTACAGGCTGACCGTGGCGGACGTATGCGAGGCTGTCTGCTGC 441
QY 621 CTTAGACATGCCAACCGGCTGTCTCTTCACTGGGTCAAGGCTTACGAGCACCG 680
DB 442 GGTGTTCTTGGCGGTGCGCATGTGACCGACGTTGGTGGGTGTTCTTACCCCGTGA 501
QY 681 AATCCGCCCCCCTGGAGTGTGG 706
DB 502 GTCCCGTCCACCGAGCTGCTGG 527

RESULT 7

CG034209

LOCUS

DEFINITION PUIG50TD ZM 0.6_1.0_KB Zea mays genomic clone ZMMBma0666103, GSS 19-AUG-2003

genomic survey sequence.

ACCESSION CG034209

VERSION CG034209.1

KEYWORDS GSS.

SOURCE Zea mays

ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD


```
REFERENCE 1 (bases 1 to 815)
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, W.A., Bedell, J.A., Ronliff, T.,
Citek, R.W., Nunberg, A., Robbins, D., and Lakey, N.
Benutzer, J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
    source
    1..815
        /organism="Zea mays"
        /mol_type="genomic DNA"
        /strain="B73"
        /db_xref="taxon:4577"
        /clone="ZM06103"
        /clone_lib="ZM 0.6 1.0 KB"
        /note="Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
        Cot selected genomic DNA library"
ORIGIN
Query Match 5.2%; Score 53.2; DB 29; Length 815;
Best Local Similarity 46.1%; Pred. No. 2.2;
Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
QY 321 CCTCAGGAGATTGCCGCTCCACATCCCTACGGCAACCGGAGCTCTGGAGGAGGT 380
DB 138 CTTTGAATGCTGTCCCGTCTTCCCGGTACCGTACGTACCTGCAACGCGGACCGGC 197
QY 381 GGGGACGCTGCTTTCATGTCCTCCCTGAGATGTTGGCCCTCAACCTGGGGGTACCCG 440
DB 198 GCTCAGGGCTGCTTCATCGGGGCTCGCGCTCTGCTGCTCTGCTCTCTCTCACCGA 257
QY 441 GCAGACCTGTCAGGCTGGAAGAGTCTTTCAGAAAGGCTGTTGGCCACCGAGCT 500
DB 258 CAGTACCGCGACACCGCCACGGGCGCTGCGGTTCAGCTTCAACACCGGCGCG 317
QY 501 CTTTCAACAAACCTCAACGGGAGCGCGGCCATCGGCACCTTTGGGCGCTCCGCT 560
DB 318 CTGTCTCCCATCGAGGTGGCGGCTCCAGCTCCGCGCCCGCGCGCGCGCGCC 377
QY 561 GAGCGCAGGAAGCAGCTCACCCTGGAACGACTACATCTACCTCGAGGAACTCGC 620
DB 378 GCGGACGACAGGTACAGCTGACCTGCGCGACGTCATGACCGGGCTGCTGTGTTGCG 437
QY 621 CCTAGACATGGCCAAAGGCTGCTCTCTTCACTGGGTCAAGGCTTACGAGGACACCG 680
DB 438 GGTGTTCTGGCGGTGGCCATAGTGGACCGAACGTTGGTGGCTGCTTCTACCCCGTGA 497
QY 681 AATCGCCCGCCCTGGAGCTGCTGG 706
DB 498 GTCCGCGTCCACGAGCAGCTGCTGG 523
RESULT 8
CC661216 897 bp DNA linear GSS 19-JUN-2003
LOCUS OGLAU18TH ZM 0.7 1.5 KB Zea mays genomic clone ZM0610312C11,
DEFINITION genomic survey sequence.
ACCESSION CC661216
VERSION CC661216.1 GI:32064967
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays (fruit fly)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 897)
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, W.A., Bedell, J.A., Ronliff, T.,
Citek, R.W., Nunberg, A., Robbins, D., and Lakey, N.
Benutzer, J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.
FEATURES
    source
    1..897
        /organism="Zea mays"
        /mol_type="genomic DNA"
        /strain="B73"
        /db_xref="taxon:4577"
        /clone="ZM0610312C11"
        /clone_lib="ZM 0.7 1.5 KB"
        /note="Vector: pBCSK-; Site 1: HindII; 0.7-1.5 kb
        methylation filtered genomic DNA library"
ORIGIN
Query Match 5.2%; Score 53.2; DB 29; Length 897;
Best Local Similarity 46.1%; Pred. No. 2.3;
Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
QY 321 CCTCAGGAGATTGCCGCTCCACATCCCTACGGCAACCGGAGCTCTGGAGGAGGT 380
DB 412 CTTTGAATGCTGTCCCGTCTTCCCGGTACCGTACGTACCTGCAACGCGGACCGGC 471
QY 381 GGGGACGCTGCTTTCATGTCCTCCCTGAGATGTTGGCCCTCAACCTGGGGGTACCCG 440
DB 472 GCTCAGGGCTGCTTCATCGGGGCTCGCGCTCTGCTGCTCTGCTCTCTCTCACCGA 531
QY 441 GCAGACCTGTCAGGCTGGAAGAGTCTTTCAGAAAGGCTGTTGGCCACCGAGCT 500
DB 532 CAGTACCGCGACACCGCCACGGGCGCTGCGGTTCATGACACCGAGCGCGC 591
QY 501 CTTTCAACAAACCTCAACGGGAGCGCGGCCATCGGCACCTTTGGGCGCTCCGCT 560
DB 592 CTTGCTCCCATCGAGGTGGCGGCTCCAGCTCCGCGCCCGCGCGCGCGCGCC 651
QY 561 GAGCGCAGGAAGCAGGCTCACCTGGAGGCTACATCTACCTGGAGGAACTCGC 620
DB 652 GCGGACGACAGGTACAGGCTGACCTGCGCGAGCTCATGACGCGGCTGCTGTGTTGCG 711
QY 621 CTTAGACATGGCCAAACGGCTGCTCTTTCATCTGAGTCAAGGCTTACGAGGACACCG 680
DB 712 GGTGTTCTGGCGGTGGCCATAGTGGACCGCAACGTTGGTGGCTGCTTCTACCCCGTGA 771
QY 681 AATCGCCCGCCCTGGAGCTGCTGG 706
DB 772 GTCCGCTCCACGAGCAGCTGCTGG 797
RESULT 9
CNS0091P 925 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TBT3 end of BAC #
DEFINITION BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053013
VERSION AL053013.1 GI:4934461
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
```



```

749 CCGTGGCGCTGACCTGGGCGCTCATCTGCTGCTCTCCCGAGGTGGAGCGTTCCAACTCC 808
|||
534 CCATGGCGGAGGATGGAGGCGCATCCCAACGACCGCTACACGGGATCTGCATATNCA 593
|||
809 CGGCGCTTATCACCTC 825
|||
594 ACGTACGCGCGGCTC 610
|||

RESULT 11
CB870888      649 bp      mRNA      linear      EST 03-JUL-2003
LOCUS        HC15J04w CH Hordeum vulgare cDNA clone HC15J04 5-PRIME, mRNA
DEFINITION   sequence.
ACCESSION    CB870888.1 GI:30072868
VERSION      CB870888
KEYWORDS     EST.
SOURCE       Hordeum vulgare
ORGANISM     Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 649)
Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and
Graner,A.
Barley ESTs from coleoptile tissue
Contact: Stein Nils
Unpublished (2003)
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@pk-gatersleben.de
Insert Length: 649 Std Error: 0.00
Plate: 15 row: J column: 4
Seq primer: T7.
Location/Qualifiers
1. 649
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="sloop"
/db_xref="GABI:549289"
/db_xref="taxon:4513"
/clone="HC15J04"
/tissue_type="coleoptile"
/dev_stage="coleoptile, 3 days old"
/lab_host="DH10B"
/clone_lib="CH"
/note="Vector: pSPORT; Site 1: SalI (5-end of cDNA);
Site 2: NotI (3-end of cDNA); Due to the cloning system
used blue/white selection for recombinants is not 100%
reliable. Average insert size is 1.3 kb."

ORIGIN
Query Match      5.0%; Score 51.2; DB 14; Length 649;
Best Local Similarity 46.3%; Pred. No. 5.2;
Matches 167; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 447 CCGTCCAGCGCTGGAAGAGTCTTGAGAAAAGGCGCTGTGGCCACCGAGCTCTTCA 506
|||
DB 227 CATCAACCGCGAGAATAGAGCTGGGCTGCGGTGCTATCGAAGGCGGCGTGGCCCA 286
|||
QY 507 CCAAAACCGTCAACGGGGAGCGCGGCCATCGGCACCCCTTTGGGCCGCTCCGCGTGAAGCC 566
|||
DB 287 CAAGATCGACTTCCGCGAGGCGCGCGCTCCCGCTCTCTGGACGCCCTCTCTCGAGACGA 346
|||
QY 567 AGGGAAGCGAGCTCACCTCGAGAGCTACTACTACCCCTGGAGAGACCTCGCCCTAGA 626
|||
DB 347 GGCCAAACCGACCTTCGACTTCGTCTTCGTGGACGCCGCAAGGACAACTACTCTCAA 406
|||
QY 627 CATGGCAACGGCGTCTCTCTCTCAACTGGGTCAAGGCTTACGAGGACCAAGGAATCCG 686
|||

```

```

DB 407 CTACCACGAGCGCTCATGAGCTCGTCAAGGTGGGCGCTCTCTCGGCTACGACAACAC 466
|||
QY 687 CCCACCCCTGGACGTGCTGCTCTCTGGGCTCAGGGGAAAAGGCTGATGCCCAACACCAA 746
|||
DB 467 CCTCTGGAACGGCTCCGCTCGTCTCCCGCGACGCCCCCATGCGCAAGTATCATCCGCTA 526
|||
QY 747 GACCGTGGCGCTTGACCTGGGCGCTCATCTGCTGCTCTCCCGAGGTGGAGCGTTCCAAACT 806
|||
DB 527 CTACCGGAGCTTCGCTCTCGACTCAACAAGGCCCTCGCCGNCAGCACCGCGCTCGAGAT 586
|||
QY 807 C 807
|||
DB 587 C 587
|||

RESULT 12
CB874255/c
LOCUS        CB874255/c
DEFINITION   sequence.
ACCESSION    CB874255
VERSION      CB874255.1 GI:30076241
KEYWORDS     EST.
SOURCE       Hordeum vulgare
ORGANISM     Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 594)
Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and
Graner,A.
Barley ESTs from coleoptile tissue
Contact: Stein Nils
Unpublished (2003)
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@pk-gatersleben.de
Insert Length: 594 Std Error: 0.00
Plate: 15 row: J column: 4
Seq primer: SP6.
Location/Qualifiers
1. 594
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="sloop"
/db_xref="GABI:552657"
/db_xref="taxon:4513"
/clone="HC15J04"
/tissue_type="coleoptile"
/dev_stage="coleoptile, 3 days old"
/lab_host="DH10B"
/clone_lib="CH"
/note="Vector: pSPORT; Site 1: SalI (5-end of cDNA);
Site 2: NotI (3-end of cDNA); Due to the cloning system
used blue/white selection for recombinants is not 100%
reliable. Average insert size is 1.3 kb."

FEATURES
source
1. 594
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="sloop"
/db_xref="GABI:552657"
/db_xref="taxon:4513"
/clone="HC15J04"
/tissue_type="coleoptile"
/dev_stage="coleoptile, 3 days old"
/lab_host="DH10B"
/clone_lib="CH"
/note="Vector: pSPORT; Site 1: SalI (5-end of cDNA);
Site 2: NotI (3-end of cDNA); Due to the cloning system
used blue/white selection for recombinants is not 100%
reliable. Average insert size is 1.3 kb."

ORIGIN
Query Match      4.9%; Score 50.6; DB 14; Length 594;
Best Local Similarity 46.3%; Pred. No. 6.7;
Matches 167; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 447 CGTCCAGCGCTGGAAGAGTCTTGAGAAAAGGCGCTGTGGCCACCGAGCTCTTCA 506
|||
DB 588 CATCAACCGCGAGAATACGAGCTGGGCTGCGGTGCTATCGAAGGCGGCGTGGCGCA 529
|||
QY 507 CCAAAACCGTCAACGGGGAGCGCGGCCATCGGCACCCCTTTGGGCCGCTCCGCGTGAAGCC 566
|||
DB 528 CAAGATCGACTTCGCGAGGCGCGGCGCTCTCTCTCGACGCGCTCTCTCGAGGACGA 469
|||

```

QY 567 AGGAAAGCAGGCTACCTGGAGCACTACATCTACCTGGAGCACTCTCGCCCTAGA 626
 Db 468 GGCAACACACGACCTTCGACTTCGCTTCCTGGAGCGCGCAAGAGCAACTTACTCTAA 409
 QY 627 CATGGCCAAAGGGGTGCTCTCTTCAACTGGGTCAAGGCTTACAGGACACCGAATCCG 686
 Db 408 CTACCAGAGCGCTCATGAAGCTCGTCAAGGTCGGGGCTCTCGGTACGACAACAC 349
 QY 687 CCCACCTGAGCTGCTGCTCTCTGGCTCAGGGAAGAGGTATGCTCCACACCAA 746
 Db 348 CCTGTGAAGCGGTTCCTGTCTCCCGCGAGCGCCCATCGCAAGTACATCGCTA 289
 QY 747 GACCGTGGCGGTGACCTGGGCTTCACTCTGCTCTCCCGAGGTGGAGGTTCCAAAT 806
 Db 288 CTACCGGAGTCTGCTCTCGACTCAACAGGCGCTCGCGCGACGAGCGGTCTGAGAT 229
 QY 807 C 807
 Db 228 C 228

RESULT 13

BI960118 637 bp mRNA linear EST 22-OCT-2001
 LOCUS HVSMEN0023E18f Hordeum vulgare rachis EST library HVCNDA0015
 DEFINITION (normal) Hordeum vulgare subsp. vulgare cDNA clone HVSME0023E18f, mRNA sequence.

ACCESSION BI960118.1 GI:16311373

VERSION EST.
 KEYWORDS Hordeum vulgare subsp. vulgare
 SOURCE Hordeum vulgare subsp. vulgare

ORGANISM Hordeum vulgare subsp. vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE 1. (bases 1 to 637)
 Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Chin,A., Begum,D.,
 Frisch,D., Atkins,M., Yu,Y., Henry,D., Palmer,M., Rambo,T.,
 Simmons,J., Oates,R. and Main,D.

Development of a genetically and physically anchored EST resource
 for barley genomics: Morex rachis cDNA library
 Unpublished (2001)
 Contact: Wing RA

Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu

Total hg bases = 579
 Seq primer: AATAACCTCACTAAAGG
 High quality sequence start: 5
 High quality sequence stop: 633.

FEATURES

source
 1..637
 /organism="Hordeum vulgare subsp. vulgare"
 /mol_type="mRNA"
 /cultivar="Morex"
 /sub_species="vulgare"
 /db_xref="taxon:112509"
 /clone="HVSME0023E18f"
 /tissue_type="Rachis"
 /lab_host="TJC121"
 /clone_lib="Hordeum vulgare rachis EST library HVCNDA0015
 (normal)"
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
 XhoI; Plants were grown at Washington State University,
 Pullman, WA in a greenhouse, the rachises were excised and
 frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close
 lab at the University of California, Riverside total RNA
 was prepared, poly(A) was purified, one primary
 unamplified cDNA library was made, and 1 million pfu were
 in vivo excised to give pBluescript SK(-) cDNA phagemids

ORIGIN

Query Match 4.9%; Score 50.6; DB 12; Length 637;
 Best Local Similarity 46.3%; Pred. No. 6.8;
 Matches 167; Conservative 0; Mismatches 194; Indels 0; Gaps 0;

QY 447 CGTCCAGCCTGGAGAGGTCTTTGAGAAAAGGGCTGTGGCCACCGAGCTCTTCA 506
 Db 133 CATCAACCGCAGAACTACGAGCTGGGGCTGCCGTGATCAGAAGCGCGGTGGCGCA 192
 QY 507 CCAACCGTCAACGGGGAGCGCGGCATCGGCACCCCTTTGGGCCCTCCGGTGAAGCC 566
 Db 193 CAAGATCGACTTCGCGAGGGCGCGGCTCCCGCTCTCGAGCGCTCTCGAGGACGA 252
 QY 567 AGGAAAGCCAGGCTCACCCTGGAGCACTATCTACCTTGGAGGAACTCGCCCTAGA 626
 Db 253 GGCCAAACCCAGGCACTTCGACTTCGTTGGAGCGCGCAAGGACAATACCTCAA 312
 QY 627 CATGGCCAAACGGGTGCTCTCTTCAACTGGGTCAAGGCTCGGCGCTCTCGGTACGACAC 686
 Db 313 CTACCAGAGCGCTCATGAAGCTCGGCGGCTCTCGGTACGACACAC 372
 QY 687 CCCACCGTGAAGCTGCTGCTCTGGGTCTAGGGGAAAAGGTGATGCCCAACACCAA 746
 Db 373 CCTCTGAAGCGGTCTCCGTGCTTCCCGCGAGCGCCCATCGCAAGTACATCCGCTA 432
 QY 747 GACCGTGGCGGTGACCTGGGCTCATCTCTGGTCTCCCGAGGTGAGAGGTTCCAAAT 806
 Db 433 CTACCGGAGTCTGCTCTCGACTCAACAGGCGCTCGCGCGACGAGCGGTCTGAGAT 492
 QY 807 C 807
 Db 493 C 493

RESULT 14

BI962267 664 bp mRNA linear EST 26-JUL-2002
 LOCUS EBro01 SQ005_F11 R root, 3 week, hydroponic grown, no treatment, cv
 DEFINITION Optic_EBro01 Hordeum vulgare subsp. vulgare cDNA clone
 EBro01 SQ005_F11 5', mRNA sequence.

ACCESSION BI962267
 VERSION BI962267.1 GI:21970739
 KEYWORDS EST.

SOURCE Hordeum vulgare subsp. vulgare
 ORGANISM Hordeum vulgare subsp. vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Hordeum.

REFERENCE 1. (bases 1 to 664)
 Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
 Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
 Development of Barley Transcriptome Resources
 Unpublished (2001)

CONTACT: Waugh R, Marshall DF
 Genome Dynamics/Computational Biology
 Scottish Crop Research Institute
 Invergowrie, Dundee, DD2 5DA, Scotland, UK

(Chin). Phagemids were plated and picked at the Clemson
 University Genomics Institute (CUGI) (Begum, Palmer, DNA
 Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
 sequencing and sequence analysis were performed at CUGI
 (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The
 sequence has been trimmed to remove vector sequence and
 contains a minimum of 100 bases of phred value 20 or
 above. For more details on library preparation and
 sequence analysis see
 http://www.genome.clemson.edu/projects/barley. To order
 this clone see http://www.genome.clemson.edu/orders/Also
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
 Genetically and physically anchored EST resources for
 barley genomics. Barley Genetics Newsletter 31:29-30.
 (http://wheat.pw.usda.gov/gppages/bgn/31/cover.html)"

Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@sari.sari.ac.uk.

Location/Qualifiers

FEATURES
source

```
1..664
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Optic"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="EBRo01.S0005.F11"
/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="root, 3 week, hydroponic grown, no treatment, cv Optic, EBR001"
/notes="Vector: pSPOR11; Site 1: Sal I; Site 2: Not I; Non-normalised library, directionally cloned into pSPOR11. Derived from roots of 3 week old hydroponically grown unstressed barley plants. Developed as part of the barley transcriptome resources of BBSRC/SEERAD funded cereal IGF (Investigating Gene Function) project."
```

ORIGIN

```
Query Match 4.9%; Score 50.6; DB 13; Length 664;
Best Local Similarity 46.3%; Pred. No. 6.9; Mismatches 0; Gaps 0;
Matches 167; Conservative 0; Indels 194; Indels 0; Gaps 0;

QY 447 GGTCCACCGCTGGAGAAGGCTCTTGAGAAAAGGGCTGTGGCCACCGACGCTCTTCA 506
Db 140 CATCAACCGGAGAACTACGAGCTGGGGCTGCGTGCATCGAGAAGGCGCGGTGGGCGCA 199
QY 507 CCAACCTGTCAACGGGAGCGCGGCCATCGGACACCTTTGGCGCTCGCGCTGAGGCC 566
Db 200 CAAGATCGACTTCGCGAGGGCGCGCGCTCCCGTCTGGACGCCCTCTCTCGAGGACGA 259
QY 567 AGGGAAGCCAGGCTCACCGCTGGACGACTACATCTACCCCTGGAGAACCTCGCCCTAGA 626
Db 260 GGCACACCGACGACCTTGGACTTCTGCTGTGGAGCGCGACAGGACAACTACCTCAA 319
QY 627 CATGGCCACGGCGTGTCTCTTCACTGGGTCAAGGCTTACAGGCTACGAGGACGGAATCG 686
Db 320 CTACACGAGCGCTCATGAAGCTCGTCAAGGTGCGCGGCTCTCTCGGCTACGACAAAC 379
QY 687 CCCCACCTTGGACGCTGTCTCTGCGCTCAGGGGAAAGGGTGTATGCCCAACACCAA 746
Db 380 CTTCTGGAACGCTCGCTGCTCTCCCGCGACGCCCGCCATGCGCAAGTATACCGCTA 439
QY 747 GACCGTGGCGTTGACCTGGGCTCATCTGTCTCTCCCGAGGTGGAGCGTTCCAACT 806
Db 440 CTACCGGACTTCTGCTCTCGACTCAACAAGGCCCTCGCGCGCCAGCGCGTCTGAGAT 499
QY 807 C 807
Db 500 C 500
```

RESULT 15

CA275829
LOCUS SCCSD1091C07.9 SD1 Saccharum officinarum cDNA clone SCCSD1091C07
DEFINITION 5', mRNA sequence.

ACCESSION CA275829

VERSION CA275829.1 GI:35994669

KEYWORDS EST.

SOURCE Saccharum officinarum

ORGANISM Saccharum officinarum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Saccharum.

REFERENCE 1 (bases 1 to 973)

AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

TITLE The libraries that made SUCEST

JOURNAL

COMMENT

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bccccenter.fcav.unesp.br
Plate: 091 row: C column: 07
Seq primer: T7 Promoter Primer.

FEATURES

source

```
1..973
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCCSD1091C07"
/lab_host="DH10B"
/clone_lib="SD1"
/notes="Organ: Developing seeds (large insert library);  
Vector: pSPOR1; Site 1: SalI; Site 2: NotI; An  
unidirectional cDNA library generated from [developing  
seeds (large insert library)]. cDNA was prepared from  
polyA+ mRNA using SuperScript plasmid System Kit  
(Invitrogen). The double-strand cDNAs were fractionated  
in a sepharose CL-2B 40cm-columns and fragments sizing  
between 0.8 and 1.5 Kb were directionally cloned into the  
vector. Details of each source of RNA and library  
construction can be obtained at  
http://sucest.lad.ic.unicamp.br/public"
```

ORIGIN

```
Query Match 4.9%; Score 50.6; DB 14; Length 973;
Best Local Similarity 46.9%; Pred. No. 7.8;
Matches 159; Conservative 0; Mismatches 179; Indels 0; Gaps 0;

QY 418 GCCTCAACCTGGGGGTACCCGGCAGACGCTCCACGCTGGGAAGGTCCTTGAGAAA 477
Db 26 GCGCAGCGCGCGCGCTGACCGCAGGAGCGCGCGCGCTCGCGCTCAGCCTCAGC 85
QY 478 AAGGGCTGTGGCCACCGACGTCCTTCAACCAACGTCACCGGAGCGCGCGGCATC 537
Db 86 AAGGCGCTGCCAAGGCGCGCGCTTCCAGGAACTCTGAGGCTCTGGCCCG 145
QY 538 GGCACCCCTTTGGGCGCTCGCGCTCAGGCCAGGAAAGCCAGGCTCACCTTGGACGACTAC 597
Db 146 CGCATCTCCCGCGCGCGAGCTCGCGCTTCCGCTCCCGCTCTCTCTCCGCAAGCTC 205
QY 598 ATCTACCCCTGGAGAACTCGCCCTTAGACATGGCCAAAGGCGTGTCTCTTCAACTGG 657
Db 206 CTGCTGGCACCCCGCAACACAGCAAGGGCGGGCGGCCAGAGCCCAAGATCGAC 265
QY 658 GTCAAGGCTTACCAGGACACCGGAATCCGCCCCCAGCTGTGACGTGTCTCTGGGCT 717
Db 266 TTCAAGACCGCGTGGACCACTTCTGCTTCCACCCCGGGGCGGGGTGATCGAGCG 325
QY 718 CAGGGGAAAAGGTGATGCCCAACACCAAGACCGGTGG 754
Db 326 GTGAGGAAGAGCTGGGCTCTCAGCGCCACGAGCTGG 362
```

Search completed: March 18, 2004, 05:45:28

Job time : 1667.36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 20:21:47 ; Search time 14869.6 Seconds
(without alignments)
17049.054 Million cell updates/sec

Title: US-09-664-186-6
Perfect score: 5849
Sequence: 1 tctagaaggtcagggtggac.....tctagaagaccagaagacc 5849

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_on.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_to.*
- 27: em_ats.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	5849	100.0	5849	1	AF013571	Thermus s
2	5849	100.0	5849	6	AR139426	Sequence
3	1600	27.4	1600	1	U89376	Thermus sp.
C 4	1242	21.2	1242	6	AR031809	Sequence
C 5	1026	17.5	1026	6	AR139424	Sequence
6	999	17.1	999	6	AR031810	Sequence
7	187.2	3.2	4938	1	AF328924	Helicobac
8	185.6	3.2	11131	1	AE001444	Helicobac
9	184	3.1	4931	1	AF328909	Helicobac
10	182.4	3.1	3172	1	AF328912	Helicobac
11	180.8	3.1	3261	1	AF328911	Helicobac
12	180.8	3.1	5001	1	AF328915	Helicobac
C 13	180	3.1	180	6	AR139425	Sequence
14	179.2	3.1	3184	1	AF328910	Helicobac
15	179.2	3.1	4931	1	AF328916	Helicobac
C 16	108	1.8	1556	1	ECCEAIMG	X17111 Enterobacte
C 17	77	1.3	7218	6	I66494	Sequence 14
C 18	61	1.0	3141	1	AF051092	Xanthomon
C 19	59.8	1.0	303040	1	AP000991	Thermopla
C 20	59.8	1.0	308050	1	SC0939124	Streptomy
C 21	59	1.0	2000	6	AX655393	Sequence
22	58.8	1.0	125020	9	AF429315	Homo sapi
23	57.6	1.0	2000	6	AX655393	Sequence
C 24	57.2	1.0	317022	1	AE017148	Helicobac
C 25	56.8	1.0	125020	9	AF429315	Homo sapi
26	56.8	1.0	308000	1	SC0939112	Streptomy
C 27	56.4	1.0	2031	6	AX653795	Sequence
C 28	56.4	1.0	2031	6	AX755741	Sequence
29	56.4	1.0	7218	6	I66494	Sequence 14
C 30	55.4	0.9	2028	6	AX660328	Sequence
C 31	55.4	0.9	2028	6	AX755777	Sequence
C 32	54.8	0.9	2028	6	AX653794	Sequence
C 33	54.8	0.9	2060	6	AX654685	Sequence
C 34	54.8	0.9	2273	8	AK106696	Oryza sat
35	54.8	0.9	137651	8	AC130600	Oryza sat
36	54.8	0.9	177374	8	AC135429	Oryza sat
C 37	54.6	0.9	202921	2	AC119314	Rattus no
C 38	54.6	0.9	299050	1	SC0939119	Streptomy
39	53.6	0.9	299425	1	AP005037	Streptomy
C 40	53.6	0.9	303550	1	SC0939118	Streptomy
C 41	53.2	0.9	1755	6	AX654077	Sequence
C 42	53.2	0.9	1965	6	AX659936	Sequence
C 43	53.2	0.9	1965	6	AX755739	Sequence
C 44	53.2	0.9	2303	8	AK072823	Oryza sat
C 45	53.2	0.9	100800	8	AC105260	Oryza sat

ALIGNMENTS

RESULT 1
AF013571
LOCUS
DEFINITION
Thermus sp. YS45 plasmid pTSP45s: Tsp45I restriction endonuclease
(tsp45IR), Tsp45I methyltransferase (tsp45IM), and replication
protein (rept) genes, complete cds.
AF013571
AF013571.1 GI:2735918
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Thermus sp. YS45
Thermus sp. YS45
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
Thermus.
REFERENCE
1 (bases 1 to 5849)
Wayne, J. and Xu, S.Y.

TITLE Identification of a thermophilic plasmid origin and its cloning within a new *Thermus*-*E. coli* shuttle vector

JOURNAL Gene 195 (2), 321-328 (1997)

MEDLINE 97449309

PUBMED 9305778

REFERENCE 2 (bases 1 to 5849)

AUTHORS Wayne, J., Holden, M. and Xu, S.Y.

TITLE The Tsp45I restriction-modification system is plasmid-borne within its thermophilic host

JOURNAL Gene 202 (1-2), 83-88 (1997)

MEDLINE 98087421

PUBMED 9427549

REFERENCE 3 (bases 1 to 5849)

AUTHORS Wayne, J., Holden, M. and Xu, S.-Y.

TITLE Direct Submission

JOURNAL Submitted (10-JUL-1997) Research, New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA

FEATURES

source

1..5849

Location/Qualifiers

/organism="Thermus sp. YS45"

/mol_type="genomic DNA"

/strain="YS45"

/db_xref="taxon:64227"

/plasmid="pisp45s"

<602..>1612

/gene="tsp45IR"

602..606

/gene="tsp45IR"

623..1612

/gene="tsp45IR"

/notes="R Tsp45I"

/codon_start=1

/transl_table=11

/product="Tsp45I restriction endonuclease"

/protein_id="AAC46043.1"

/db_xref="GI:2735919"

translation="MAEMNVQTSRVELLEKGYLDKLVYKGGSSRSRSPVEEVEK
LREAYKAYGRQDSPEATKLVNARKVVERPFPHPYLPVYLVSEKAEKANK
ALEEALQEVKHPETIRLKEAQRGVEALIQRLKPPENRQIGMFKWYKBEK
KGIIEBLPCPKPKTVVSPKSPKPEQAPLIAERAGIIYTGDEALKDAKENLG
LGEAEBLGTGVDFYVIRSPETWHLTGEVYKFSQFGNQDNOKLVAKASIRLDLE
KRHIGTVVDMGVSKFSGWAGLGRKETTIVTSVLLLPDLIAELYKGEALGL"

complement (1609..>2860)

/gene="tsp45IM"

complement (1609..2850)

/gene="tsp45IM"

/notes="M Tsp45I"

/codon_start=1

/transl_table=11

/product="Tsp45I methyltransferase"

/protein_id="AAC46044.1"

/db_xref="GI:2735920"

translation="MSRSPGLTRKAPLKASKTSBPSPRLVYPKRDEKEILDQPTP
QLVIRKETLFLGGNAPLPEIDPIGTFLGNGQVLRWMLREFPGYAGKVQVLYIDPP
YGTGQQLVGGDETRVATVSQPKNGQLGDDTLDGPQVFEFLRLLILLRELWADSG
LTFVHDEKASVEVKILDEIVFRNFMVHAIASNPKNFSKAFGSKOMILVYSK
TRDYVNESAPVSEIEARLAPFVDNGERVTTNPLHAGETKDGTPGRWRGLPP
PGRWRYPEKIDELDAQGLVNSKNGVRKKVARDRLKKGKLDVWQKDPVPR
YPERNLMKLIVTQSGNEGDLVDFPAGSGTTLIASPLLRKRSIGIDASWEAVKAF
TRVLEDFPRLQKHFIVSAF"

2164..4500

/notes="required for plasmid replication (putative)"

complement (2856..2860)

/gene="tsp45IM"

complement (<3132..>4169)

/gene="repT"

complement (3132..4157)

/gene="repT"

/function="required for plasmid replication"

/notes="RepT"

/codon_start=1

/transl_table=11

/product="replication protein"

/protein_id="AAC46045.1"

/db_xref="GI:2735921"

translation="MKNEKTFREELVEALEETHDNTATRGSDRGSEDFFLATDPPPD
GGAENLAKGFTYQKALRIALPEKDHAPLSVSGAPPPIPPAEPVGVNVCQAVDGPQ
KLJELLOEIVARSITIPYGNRELNRKVGTVVFMVPLEMLANLGVTRQTVHAKVLEKK
GLVATDVHQTVNGERRAIGTLWALRFGKARLTDLDYIYPWNLALDMANGVLSFN
WVAYQDHGIRPTLDVLVLAQKRVNPKTKVADGLILILPEVERSKIPALITLI
ATVIADLLDRSRFRFYAGLLMAVARGELPAQYLFAVLMRVIRDTYDGHLTREPGAYLV
KTLKEAS"

complement (3200..3208)

/misc_feature

/gene="repT"

/notes="putative DnaA box required for plasmid replication"

complement (3335..3343)

/misc_feature

/gene="repT"

/notes="putative DnaA box required for plasmid replication"

complement (4165..4169)

/RBS

/gene="repT"

complement (4194..4199)

-10_signal

/notes="putative"

complement (4218..4223)

-35_signal

/notes="putative"

complement (4253..4258)

-10_signal

/notes="putative"

complement (4276..4281)

-35_signal

/notes="putative"

ORIGIN

Query Match 100.0%; Score 5849; DB 1; Length 5849;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 5849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAGAAGGTTCAGGTGCGACAGGAAACACCATAGCCCTGCCAAGAGATCGAGGT 60

DB 1 TCTAGAAGGTTCAGGTGCGACAGGAAACACCATAGCCCTGCCAAGAGATCGAGGT 60

QY 61 TGGTGTCCGAAAGTGCCTATCGGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA 120

DB 61 TGGTGTCCGAAAGTGCCTATCGGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA 120

QY 121 CCGCATTTGGCCACGAGGTTCGAGTTCGAGTTCAGCGCCACAAAGGGTCTCTCTCA 180

DB 121 CCGCATTTGGCCACGAGGTTCGAGTTCGAGTTCAGCGCCACAAAGGGTCTCTCTCA 180

QY 181 AACTTTCTTTCTAGTCCGCTTGGACGAGGCGAGGAGGAAAGGCTTCATGGGCTCA 240

DB 181 AACTTTCTTTCTAGTCCGCTTGGACGAGGCGAGGAGGAAAGGCTTCATGGGCTCA 240

QY 241 CCTCTTTCCCTCTCTCTTGGCGGCTTACGCGGCTTAAACTCTGAGACGGCTCAAGTT 300

DB 241 CCTCTTTCCCTCTCTCTTGGCGGCTTACGCGGCTTAAACTCTGAGACGGCTCAAGTT 300

QY 301 TAGGATTTTCGCTTTTCGGGATTAAGATCCGCGGCTCAGGGATCCGGATCCGCTTA 360

DB 301 TAGGATTTTCGCTTTTCGGGATTAAGATCCGCGGCTCAGGGATCCGGATCCGCTTA 360

QY 361 TCCTTCCGCTCCCTTATCTACTTCGTAATGTGTCCTTGGTACTTTAAACCGTCTGAAA 420

DB 361 TCCTTCCGCTCCCTTATCTACTTCGTAATGTGTCCTTGGTACTTTAAACCGTCTGAAA 420

QY 421 CTCTCTTAACAGAGACGACAAACCTCTAAACCTTCAATCCACCGATCCAGTATA 480

DB 421 CTCTCTTAACAGAGACGACAAACCTCTAAACCTTCAATCCACCGATCCAGTATA 480

QY 481 CCATAAATGGCACAAGTTTTTGAAGAGTGGTCAAAACAAAGGCTTTCTCGGTCAAGTT 540

DB 481 CCATAAATGGCACAAGTTTTTGAAGAGTGGTCAAAACAAAGGCTTTCTCGGTCAAGTT 540

QY 541 ATGTGAGGTGGGGCGGTCAAAAGCGGACTTAAGTTTGGTAAAGCGGGAGAACAAA 600

DB 541 ATGTGAGGTGGGGCGGTCAAAAGCGGACTTAAGTTTGGTAAAGCGGGAGAACAAA 600

QY 601 CCGGGGTGTTACCATGCAACAGATGGCCGAGTGGACGCTGTGGACACAGAGAGCGTTGA 660

Db 601 CCGGGTGTACATGCAACAGATGGCGAGTGGAACGTGTGGACACAGAGAACGGTTGA 660
Qy 661 GCTTCTGAGAGGGGTATTTGGATAAACTACTCGAGGTCTATATAAGGGGAAAGTGGCTC 720
Db 661 GCTTCTGAGAGGGGTATTTGGATAAACTACTCGAGGTCTATATAAGGGGAAAGTGGCTC 720
Qy 721 TTCAGGTCAGTACGAGAGGAGTAGAGGAAAACTTCGCGAGGCCCTACAGGCATACGA 780
Db 721 TTCAGGTCAGTACGAGAGGAGTAGAGGAAAACTTCGCGAGGCCCTACAGGCATACGA 780
Qy 781 GGGGAGGAGGATAGTCCGGAGGAGAAAGAAACTCGTGAAGCCGTGCTAAATGCCAG 840
Db 781 GGGGAGGAGGATAGTCCGGAGGAGAAAGAAACTCGTGAAGCCGTGCTAAATGCCAG 840
Qy 841 AAAAAGGTGAGCGGTCCCTTCAATCAACCCCTACTCGCTTTGGTCTACTACCTGGT 900
Db 841 AAAAAGGTGAGCGGTCCCTTCAATCAACCCCTACTCGCTTTGGTCTACTACCTGGT 900
Qy 901 TTCGAAAAAGCAGAAAAAGCGAAACAGGCCCTTGGAGGAGCATTTGAGAGGTTGGCTC 960
Db 901 TTCGAAAAAGCAGAAAAAGCGAAACAGGCCCTTGGAGGAGCATTTGAGAGGTTGGCTC 960
Qy 961 AAAGCACCAGAAACCATCCCGTCTTGCCCAAGAAAGCGCAAGAGCGGTAGAAC 1020
Db 961 AAAGCACCAGAAACCATCCCGTCTTGCCCAAGAAAGCGCAAGAGCGGTAGAAC 1020
Qy 1021 CTTGATCCAAAGGCTCAGAGGACCTCCCGAATAAATCGGCAGATAGCGCGATTTCAA 1080
Db 1021 CTTGATCCAAAGGCTCAGAGGACCTCCCGAATAAATCGGCAGATAGCGCGATTTCAA 1080
Qy 1081 AAGGTGGTACAAAGAGAGCTAAAGGGGAAATAGAGAGAGGTTCCAGGCCCTACCAA 1140
Db 1081 AAGGTGGTACAAAGAGAGCTAAAGGGGAAATAGAGAGAGGTTCCAGGCCCTACCAA 1140
Qy 1141 ACCAAAGATTGTGTAGTATCCCTGAAAAAAGTAAACCGGAGCAAGCAACCCCTATTGC 1200
Db 1141 ACCAAAGATTGTGTAGTATCCCTGAAAAAAGTAAACCGGAGCAAGCAACCCCTATTGC 1200
Qy 1201 GGAGAGAGAGCGGSCATCATATACACGGGATCGGATCAAGCTTTGAAGATGCCG 1260
Db 1201 GGAGAGAGAGCGGSCATCATATACACGGGATCGGATCAAGCTTTGAAGATGCCG 1260
Qy 1261 CAAGAAAACTGGCCCTTGGCGAGGAGAGCACTAGGCAACCAAGGCGTAGATTCTA 1320
Db 1261 CAAGAAAACTGGCCCTTGGCGAGGAGAGCACTAGGCAACCAAGGCGTAGATTCTA 1320
Qy 1321 CGTGTCTATCGGCGTAGCTGAGAGACATGGCACCCTACAGGAGAGTGAAGTTTCA 1380
Db 1321 CGTGTCTATCGGCGTAGCTGAGAGACATGGCACCCTACAGGAGAGTGAAGTTTCA 1380
Qy 1381 ATCCGACTTTGGCGAAACCAAGCAACCAAGAACTAGTAGCAAGGCTTCCATAAGGTT 1440
Db 1381 ATCCGACTTTGGCGAAACCAAGCAACCAAGAACTAGTAGCAAGGCTTCCATAAGGTT 1440
Qy 1441 GGACCTTGAGAGAGGACATAGGATAGTGTGTGTGACCGGATGCTGTGTGTGAGCAA 1500
Db 1441 GGACCTTGAGAGAGGACATAGGATAGTGTGTGTGACCGGATGCTGTGTGTGAGCAA 1500
Qy 1501 GTTTCGTGGTGGGCGGACTGGGAGAAAGAAACGATCGTTACATCCGTAATCTCTCC 1560
Db 1501 GTTTCGTGGTGGGCGGACTGGGAGAAAGAAACGATCGTTACATCCGTAATCTCTCC 1560
Qy 1561 AGACCTGATAGCGAGCTCTACCAAAAGGGTCAAGAGCCCTGGGCGCTCTAGAGGGGGA 1620
Db 1561 AGACCTGATAGCGAGCTCTACCAAAAGGGTCAAGAGCCCTGGGCGCTCTAGAGGGGGA 1620
Qy 1621 CACAATCTCAAACTGT 1680
Db 1621 CACAATCTCAAACTGT 1680
Qy 1681 TTTGACCGCCTCCAGGAGGATCTATGCGGATGATCGCGCTTTAAGAGGGGTGAGGC 1740
Db 1681 TTTGACCGCCTCCAGGAGGATCTATGCGGATGATCGCGCTTTAAGAGGGGTGAGGC 1740

Qy 1741 TATAAGCGTAGTACCGGAGCCTCGAGAGGATCGAGCACTAAATCCCCCTCGTTACTCCC 1800
Db 1741 TATAAGCGTAGTACCGGAGCCTCGAGAGGATCGAGCACTAAATCCCCCTCGTTACTCCC 1800
Qy 1801 TGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTCTCGTGGGTATCGCGGGTACGG 1860
Db 1801 TGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTCTCGTGGGTATCGCGGGTACGG 1860
Qy 1861 AGGATCCTTTGAATGCCAAACGTCTGGAGCTTCTTCCCTTCTTTCAGGCGATCCGAGC 1920
Db 1861 AGGATCCTTTGAATGCCAAACGTCTGGAGCTTCTTCCCTTCTTTCAGGCGATCCGAGC 1920
Qy 1921 GTAAATCTTTTCCCGCGCACCCCGTCTTTCACAGCAATAAGCCCTTGAGCGTCTAG 1980
Db 1921 GTAAATCTTTTCCCGCGCACCCCGTCTTTCACAGCAATAAGCCCTTGAGCGTCTAG 1980
Qy 1981 CTCGTCAAGCTTCTTCCCGGGGATAGCGCAATGCGCTCCAGGAGGGGAAGTATTCCTCG 2040
Db 1981 CTCGTCAAGCTTCTTCCCGGGGATAGCGCAATGCGCTCCAGGAGGGGAAGTATTCCTCG 2040
Qy 2041 CMAGGCCCTTCGCTAGGCGCATCTTGTCTTTCAGAGGATCGAGGGGATTTGGTGT 2100
Db 2041 CMAGGCCCTTCGCTAGGCGCATCTTGTCTTTCAGAGGATCGAGGGGATTTGGTGT 2100
Qy 2101 GTACCGTTCCTCGTCTACAAAGGGGAAAGCCCTAGCGATCTCTCTTCGGAATA 2160
Db 2101 GTACCGTTCCTCGTCTACAAAGGGGAAAGCCCTAGCGATCTCTCTTCGGAATA 2160
Qy 2161 GGGCTAGCGGATTCGTTCCAAACGTAGTCCCGGTTTTGGAGTAGACAGGATCATGTC 2220
Db 2161 GGGCTAGCGGATTCGTTCCAAACGTAGTCCCGGTTTTGGAGTAGACAGGATCATGTC 2220
Qy 2221 CTTTTCGGATCCGAGGCCCTTACGGGAAAGTTTTGGATTTGAAGCATGCGGCGAT 2280
Db 2221 CTTTTCGGATCCGAGGCCCTTACGGGAAAGTTTTGGATTTGAAGCATGCGGCGAT 2280
Qy 2281 ATGTTTAAAGAAAGTTTCGCGGCCAAAGACCTCATCAAGGATGAGCTTCACTCGAACCC 2340
Db 2281 ATGTTTAAAGAAAGTTTCGCGGCCAAAGACCTCATCAAGGATGAGCTTCACTCGAACCC 2340
Qy 2341 GTATTTCTGTCTATGTGAACGAGATCAGTCTGTAGTCCGCCATCAGTCTCCTGAGAAG 2400
Db 2341 GTATTTCTGTCTATGTGAACGAGATCAGTCTGTAGTCCGCCATCAGTCTCCTGAGAAG 2400
Qy 2401 TATCAAGCGCTCCCTCAGGAACTCCACAACTGAGGACCATCGAGGTTGTATCGTAGCC 2460
Db 2401 TATCAAGCGCTCCCTCAGGAACTCCACAACTGAGGACCATCGAGGTTGTATCGTAGCC 2460
Qy 2461 CAATGACCGTTTTTGGGTGGCTGACGCGATCTGTTTCATCGCGCCAAAC 2520
Db 2461 CAATGACCGTTTTTGGGTGGCTGACGCGATCTGTTTCATCGCGCCAAAC 2520
Qy 2521 GAGAACTGTGCGCGTTTCCATAAGCGGGTCAATATAGCACTGACCTTCCCGCG 2580
Db 2521 GAGAACTGTGCGCGTTTCCATAAGCGGGTCAATATAGCACTGACCTTCCCGCG 2580
Qy 2581 ATACCACACAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTTCCCCAAAAAGTA 2640
Db 2581 ATACCACACAGGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTTCCCCAAAAAGTA 2640
Qy 2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGCATTTCCCTTAGGAGGAGGGGTTTC 2700
Db 2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGCATTTCCCTTAGGAGGAGGGGTTTC 2700
Qy 2701 TTTTCGAAAAAAGTTGTGGGTGGGCTGATCAAGAACTCTCCTCTCATCGGTTTTTC 2760
Db 2701 TTTTCGAAAAAAGTTGTGGGTGGGCTGATCAAGAACTCTCCTCTCATCGGTTTTTC 2760
Qy 2761 GGGGTAGACCAACTTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGCTTT 2820
Db 2761 GGGGTAGACCAACTTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGCTTT 2820

QY 2821 TCGGTCACACAGGGTAGTACGGCTCATTTTCCTCCACACAGCGCTCTTAAGCAG 2880
Db 2821 TCGGTCACACAGGGTAGTACGGCTCATTTTCCTCCACACAGCGCTCTTAAGCAG 2880
QY 2881 ACCTCATACCCCAACACCTCAGCACTCAACCAAGGAATCGCCAAAGGGCGCTTACC 2940
Db 2881 ACCTCATACCCCAACACCTCAGCACTCAACCAAGGAATCGCCAAAGGGCGCTTACC 2940
QY 2941 TTTTGAAGCCGATATCTTCCCTGACGTATAGACCTTCGGATCGCTCAAGGCTGCAACCGA 3000
Db 2941 TTTTGAAGCCGATATCTTCCCTGACGTATAGACCTTCGGATCGCTCAAGGCTGCAACCGA 3000
QY 3001 AGATGTCGCAAGCTCTCGGGGTCAGTACACGGGCTTATCCCTCATGACACAACCT 3060
Db 3001 AGATGTCGCAAGCTCTCGGGGTCAGTACACGGGCTTATCCCTCATGACACAACCT 3060
QY 3061 TACCCCAAGAGGACACATCAATATGGGCAAGTAGACAACGAGACCAAAAGTT 3120
Db 3061 TACCCCAAGAGGACACATCAATATGGGCAAGTAGACAACGAGACCAAAAGTT 3120
QY 3121 GGGCCACTCTCAGGAGGCTCTTGAGGCTCTTCACTAGGTAGCTCCCGGTCTGTC 3180
Db 3121 GGGCCACTCTCAGGAGGCTCTTGAGGCTCTTCACTAGGTAGCTCCCGGTCTGTC 3180
QY 3181 AGATGGCCATCCCGTGAATCTCGGATAACCCCGATTAGGACGGCAATAGATATTCGCG 3240
Db 3181 AGATGGCCATCCCGTGAATCTCGGATAACCCCGATTAGGACGGCAATAGATATTCGCG 3240
QY 3241 GGGAGTTCACCCCTGGCCACAGCCACAGGAAGCTTGATAGAAAGTCTTGAACGAG 3300
Db 3241 GGGAGTTCACCCCTGGCCACAGCCACAGGAAGCTTGATAGAAAGTCTTGAACGAG 3300
QY 3301 TCATCTAGGAGATCGCAATGATAGCAATGAGGCTGATAGGAGTCTTGAAGGAGTTGGA 3360
Db 3301 TCATCTAGGAGATCGCAATGATAGCAATGAGGCTGATAGGAGTCTTGAAGGAGTTGGA 3360
QY 3361 CGTCCACTCGGGAGGACAGGATGAGGCCAGGTCAAGCCACGCTCTTGGTGTG 3420
Db 3361 CGTCCACTCGGGAGGACAGGATGAGGCCAGGTCAAGCCACGCTCTTGGTGTG 3420
QY 3421 GGCATACACCTTTTCCCTGAGCCACAGAGGACAGCAAGCTTCAAGGCTGGGGCGGATTCG 3480
Db 3421 GGCATACACCTTTTCCCTGAGCCACAGAGGACAGCAAGCTTCAAGGCTGGGGCGGATTCG 3480
QY 3481 TGGTCTGGTAGGCTTGACCCAGTTGAAGGAGAGCACCCGCTTGGCCATGCTAGGGCG 3540
Db 3481 TGGTCTGGTAGGCTTGACCCAGTTGAAGGAGAGCACCCGCTTGGCCATGCTAGGGCG 3540
QY 3541 AGGTTCTCCAGGGTAGATGATGTCGTCAGGGTAGCTTCCCTGGCTCAGC 3600
Db 3541 AGGTTCTCCAGGGTAGATGATGTCGTCAGGGTAGCTTCCCTGGCTCAGC 3600
QY 3601 CGGACCGCCCAAGGGTGGCGATGGGCCCGCGCTCCCGCTTGACGGTTGGTGAAGGAGC 3660
Db 3601 CGGACCGCCCAAGGGTGGCGATGGGCCCGCGCTCCCGCTTGACGGTTGGTGAAGGAGC 3660
QY 3661 TCGGTGGCACACAGGCTTTTCTCAAGGACCTTCTCAGGCGTGGACGCTTGGCGG 3720
Db 3661 TCGGTGGCACACAGGCTTTTCTCAAGGACCTTCTCAGGCGTGGACGCTTGGCGG 3720
QY 3721 GTGACCCCAAGGTTGAGGGCCACATCTCCAGGGGACCATGAAGACGACCTGCCACC 3780
Db 3721 GTGACCCCAAGGTTGAGGGCCACATCTCCAGGGGACCATGAAGACGACCTGCCACC 3780
QY 3781 TTCTCCAGAGCTCCCGTTGCGGTAGGGATGGTGGAGCGGCAATCTCTCGAGGAGT 3840
Db 3781 TTCTCCAGAGCTCCCGTTGCGGTAGGGATGGTGGAGCGGCAATCTCTCGAGGAGT 3840
QY 3841 TCCAGAGGCTTTCAGGCGGCTTGGACGCTTGGACATACATTCACCAACGGGGGTTCA 3900
Db 3841 TCCAGAGGCTTTCAGGCGGCTTGGACGCTTGGACATACATTCACCAACGGGGGTTCA 3900
QY 3901 GCTGGTGGTAGGGGGGGCCCAACAGAGGAAAGGAAGCCTCATGTCCTTCGGGT 3960

Db 3901 GCTGGTGGTAGGGGGGGCCCAACAGAGGAAAGGAAGCCTCATGTCCTTTCGGGT 3960
QY 3961 AAAGCAATCTAAGTGTCTCTTTTGGTATGTAAAGCCCTTCGCGAGCGGATTTTCGGCA 4020
Db 3961 AAAGCAATCTAAGTGTCTCTTTTGGTATGTAAAGCCCTTCGCGAGCGGATTTTCGGCA 4020
QY 4021 CCTCCATCTCGAGGGGGTGGCCAAAGAAAGTCTCTGACCCCTCATCTGACCCC 4080
Db 4021 CCTCCATCTCGAGGGGGTGGTGGCCAAAGAAAGTCTCTGACCCCTCATCTGACCCC 4080
QY 4081 CTAGTGCAATCGGTTGTCTGGGTTTCTCTAAAGCCCTCGTAAAGCTCTTCAAAAG 4140
Db 4081 CTAGTGCAATCGGTTGTCTGGGTTTCTCTAAAGCCCTCGTAAAGCTCTTCAAAAG 4140
QY 4141 GTTTTTTCGTTCTTCACCCCTCGGACCTCTTGTGATCTGAGCCCGGAGGCTTACCCCTAG 4200
Db 4141 GTTTTTTCGTTCTTCACCCCTCGGACCTCTTGTGATCTGAGCCCGGAGGCTTACCCCTAG 4200
QY 4201 GTCTTGGGGGTGATCCGGGGCAACCGCTCGGTTTGGCCCTTTTATGGGTCCAAAATAAC 4260
Db 4201 GTCTTGGGGGTGATCCGGGGCAACCGCTCGGTTTGGCCCTTTTATGGGTCCAAAATAAC 4260
QY 4261 CGTCAGCCCAAGCGCTGGCAATCCCCCTCTCTAAAGCCCGTTATAGGCCCTCTTAGGAG 4320
Db 4261 CGTCAGCCCAAGCGCTGGCAATCCCCCTCTCTAAAGCCCGTTATAGGCCCTCTTAGGAG 4320
QY 4321 GGGGGTAGTACTTTCTTACCCCTTAGGCTTGGAGAGGCTTAGGAGGCTCTCTAGGGCC 4380
Db 4321 GGGGGTAGTACTTTCTTACCCCTTAGGCTTGGAGAGGCTTAGGAGGCTCTCTAGGGCC 4380
QY 4381 TCGTGGGGGTGAGGGGTAACTCATGGCCAGCCCGCGGCTCGGACTCTCGAGGAGG 4440
Db 4381 TCGTGGGGGTGAGGGGTAACTCATGGCCAGCCCGCGGCTCGGACTCTCGAGGAGG 4440
QY 4441 CCTCCATAGCCTACTCGTGTGGAGTTTGTGAAGGGTTCACTAATGATACGGCTAGC 4500
Db 4441 CCTCCATAGCCTACTCGTGTGGAGTTTGTGAAGGGTTCACTAATGATACGGCTAGC 4500
QY 4501 CTCGGGATCAGGCAAAATGATGAGGTTTGGTATAAACCCTCAGGTTTGAAGCTA 4560
Db 4501 CTCGGGATCAGGCAAAATGATGAGGTTTGGTATAAACCCTCAGGTTTGAAGCTA 4560
QY 4561 GTTTATGTCGGTTTATGACCTTGTGATCAGGATCAGGGCATAAACACAGTTTCTGTC 4620
Db 4561 GTTTATGTCGGTTTATGACCTTGTGATCAGGATCAGGGCATAAACACAGTTTCTGTC 4620
QY 4621 ACAGAAAGAACTTTCGGATCTAAGAGGGGAAAGAGGTGTAGAGGACGGCTTTCATG 4680
Db 4621 ACAGAAAGAACTTTCGGATCTAAGAGGGGAAAGAGGTGTAGAGGACGGCTTTCATG 4680
QY 4681 AAAGTTGGCTCTTGAAGGCGGTTGTAGAGGCGGCTCTCGGTTCAAAATCTTCTTCTC 4740
Db 4681 AAAGTTGGCTCTTGAAGGCGGTTGTAGAGGCGGCTCTCGGTTCAAAATCTTCTTCTC 4740
QY 4741 TCTCTCAGGTTTCCGAGGTTTCAGGTTCTTGGTCCAGGCTTGTGACCAAGTTTGAACA 4800
Db 4741 TCTCTCAGGTTTCCGAGGTTTCAGGTTCTTGGTCCAGGCTTGTGACCAAGTTTGAACA 4800
QY 4801 AAGTCTATTCTCGGAATATAGGGGTATCTTGTCTATCTTCTTCCCTACGGGATATCTGTCT 4860
Db 4801 AAGTCTATTCTCGGAATATAGGGGTATCTTGTCTATCTTCTTCCCTACGGGATATCTGTCT 4860
QY 4861 GTGTGAATCTGATCCCAATACATATCTCAATCTCTCAATCTCTCTCTCTCTCTCCAG 4920
Db 4861 GTGTGAATCTGATCCCAATACATATCTCAATCTCTCAATCTCTCTCTCTCTCTCCAG 4920
QY 4921 ATCCCTAATCTCTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4980
Db 4921 ATCCCTAATCTCTCTTCT 4980
QY 4981 CGACCAAGAGGCTCTCGGGGTGAGTTTCGGTAACTCGGGACAGGTTTTCATGCTCT 5040

4981	CGACCAAGACGAGCTTCTCGGGTCAGTTTCGGTAATCTCGGGACAGGTTTTTCATCGTCT	5040
5041	AGGACGAGGATTAGGCATCAAAAATGGGCTTTGACAAAATCTTCTAAAAATACTCC	5100
5041	AGGACGAGGATTAGGCATGAAAAATGGGCTTTGACAAAATCTTCTAAAAATACTCC	5100
5101	CGAGTTGGGGAAGTCCCTCGGGGAGAGATTTTGGCAGTTTAGATGTTATGCTCTAT	5160
5101	CGAGTTGGGGAAGTCCCTCGGGGAGAGATTTTGGCAGTTTAGATGTTATGCTCTAT	5160
5161	CACGGCCGGAGGCCCTCCAGATAAGTTGTCTTTGGCCCAAGTACCGGCCAGGTCGGGGT	5220
5161	CACGGCCGGAGGCCCTCCAGATAAGTTGTCTTTGGCCCAAGTACCGGCCAGGTCGGGGT	5220
5221	GCTCTTCAGGTTGGTAGTGTACTTTTACGGAAAGTTCAAAAGTCCCTTTAGAGGCTTCAG	5280
5221	GCTCTTCAGGTTGGTAGTGTACTTTTACGGAAAGTTCAAAAGTCCCTTTAGAGGCTTCAG	5280
5281	GTCCGGGATAGTCTCAAGTACTCCCAAGCGTTCTCGGGCCCGTGTGCGGGAGAGGAC	5340
5281	GTCCGGGATAGTCTCAAGTACTCCCAAGCGTTCTCGGGCCCGTGTGCGGGAGAGGAC	5340
5341	AAAGGGGTCCGGCAAAAGTTTCATCTTTGTACTTTAGGACGGATTACTTAGCACTCATAA	5400
5341	AAAGGGGTCCGGCAAAAGTTTCATCTTTGTACTTTAGGACGGATTACTTAGCACTCATAA	5400
5401	CTTCAGGGCCGTTAAGAAAGGCTCACTTCGGAGAGCGGTTGGAAGGAGCGTGGCGCTG	5460
5401	CTTCAGGGCCGTTAAGAAAGGCTCACTTCGGAGAGCGGTTGGAAGGAGCGTGGCGCTG	5460
5461	GAAAGAGACGAACCCCGATTTTTGGGAAGTCTCCCTCCAGTTTGATGATGAACGTTGGGA	5520
5461	GAAAGAGACGAACCCCGATTTTTGGGAAGTCTCCCTCCAGTTTGATGATGAACGTTGGGA	5520
5521	GGAAGCGGCACAGATGTCTTTTCAATCGCGCTCGAACCTCGGAGTACAAAGTCCTCAA	5580
5521	GGAAGCGGCACAGATGTCTTTTCAATCGCGCTCGAACCTCGGAGTACAAAGTCCTCAA	5580
5581	TGTTTCTCAGGCGCAAGAGTCTATGATGAGGTAAACCTTCGGAGTACAAAGTCCTCAA	5640
5581	TGTTTCTCAGGCGCAAGAGTCTATGATGAGGTAAACCTTCGGAGTACAAAGTCCTCAA	5640
5641	GCOCCTTTCCCAACCGCTCAAAACTCTAGGTCAGGTGTTTAGTTTTCTGAAAACT	5700
5641	GCOCCTTTCCCAACCGCTCAAAACTCTAGGTCAGGTGTTTAGTTTTCTGAAAACT	5700
5701	CTAGCTTTTCAGTGGTCAATTCCTCACCCCTCTAGCAGTACTCTGGAGTAAACCTTTG	5760
5701	CTAGCTTTTCAGTGGTCAATTCCTCACCCCTCTAGCAGTACTCTGGAGTAAACCTTTG	5760
5761	ACACAGCGGCCAAGTCTAGCGTCTCCCAAGTCCAGTTGGTCTGGGACCGCTGAGAGGGGA	5820
5761	ACACAGCGGCCAAGTCTAGCGTCTCCCAAGTCCAGTTGGTCTGGGACCGCTGAGAGGGGA	5820
5821	GGGGCTTTGGTGTAGAGGACCAAGACCC	5849
5821	GGGGCTTTGGTGTAGAGGACCAAGACCC	5849

[illegible]

Db 901 TTCCGAAAAACGAAAAACGAAACGAGCCCTTGAGAGGCGATGCGAGAGGTTGCCTC 960
Qy 961 AAAGCACCAGAAACCATCCGCTCTGCGCAAGGAAGCGCAAGAGAGGCGTAGAGC 1020
Db 961 AAAGCACCAGAAACCATCCGCTCTGCGCAAGGAAGCGCAAGAGAGGCGTAGAGC 1020
Qy 1021 CTTGATCCAAAGGCTCAAGAGAGCTCCCGAATPAATCGCGAGATAGGCGCCGATGTTCAA 1080
Db 1021 CTTGATCCAAAGGCTCAAGAGAGCTCCCGAATPAATCGCGAGATAGGCGCCGATGTTCAA 1080
Qy 1081 AAGGTGTACAAAGAGAGCTAAAGGGGAAATAGAGAGAGGCTTCCAGGCGCTACCAA 1140
Db 1081 AAGGTGTACAAAGAGAGCTAAAGGGGAAATAGAGAGAGGCTTCCAGGCGCTACCAA 1140
Qy 1141 ACCAAAGATTGTGTAGTATCCCTGAAAAAAAGTAAACCGGAGCAAGCACCCCTATTTCG 1200
Db 1141 ACCAAAGATTGTGTAGTATCCCTGAAAAAAAGTAAACCGGAGCAAGCACCCCTATTTCG 1200
Qy 1201 GGAGAGAGAGCGGGCATCATATATACAGGGATCGGATGAGAGCTTTGAAAGATGCCGC 1260
Db 1201 GGAGAGAGAGCGGGCATCATATATACAGGGATCGGATGAGAGCTTTGAAAGATGCCGC 1260
Qy 1261 CAAGGAAACCTCGGCTTGGCGAGGAAGCAGAACTAGGCAACCAAGGGCGTAGATTCTA 1320
Db 1261 CAAGGAAACCTCGGCTTGGCGAGGAAGCAGAACTAGGCAACCAAGGGCGTAGATTCTA 1320
Qy 1321 CGTGTATCCGGCTAGCGCTGAAGAGATATGGCACCTTAACAGGAGAAAGTGAATTTCA 1380
Db 1321 CGTGTATCCGGCTAGCGCTGAAGAGATATGGCACCTTAACAGGAGAAAGTGAATTTCA 1380
Qy 1381 ATCCGACTTGGCGGAAACCAAGACACAGAACTAGTAGCAAGAGGCTTCCATAGGTT 1440
Db 1381 ATCCGACTTGGCGGAAACCAAGACACAGAACTAGTAGCAAGAGGCTTCCATAGGTT 1440
Qy 1441 GGACCTTTGAGAGAGGACATAGGAATAGTGTGTGGAGCGGAATGCCCTGTGGTAGCAA 1500
Db 1441 GGACCTTTGAGAGAGGACATAGGAATAGTGTGTGGAGCGGAATGCCCTGTGGTAGCAA 1500
Qy 1501 GTTTCGTGGTGGCGGAGCTATGCGGATGGAATCGGCTTTAAGAGGGGTGAGGC 1560
Db 1501 GTTTCGTGGTGGCGGAGCTATGCGGATGGAATCGGCTTTAAGAGGGGTGAGGC 1560
Qy 1561 AGACCTGTAGAGCGAGCTTACCAAGGGTGAAGAGCCCTGGGCTCTAGAGGGCGGA 1620
Db 1561 AGACCTGTAGAGCGAGCTTACCAAGGGTGAAGAGCCCTGGGCTCTAGAGGGCGGA 1620
Qy 1621 CACAATCTCAAACTTGTGTAGCTCGGGAATTCCTTAACACCTTTCTAGTGAAGGC 1680
Db 1621 CACAATCTCAAACTTGTGTAGCTCGGGAATTCCTTAACACCTTTCTAGTGAAGGC 1680
Qy 1681 TTTGACCGCTCCAGAGGCGATCTATGCGGATGGAATCGGCTTTAAGAGGGGTGAGGC 1740
Db 1681 TTTGACCGCTCCAGAGGCGATCTATGCGGATGGAATCGGCTTTAAGAGGGGTGAGGC 1740
Qy 1741 TATAAGCGTAGTACCGAGGCTCGGAAGGATCGAGCAATAATCCCTCTGTTACTCCC 1800
Db 1741 TATAAGCGTAGTACCGAGGCTCGGAAGGATCGAGCAATAATCCCTCTGTTACTCCC 1800
Qy 1801 TGTTCGAGATGAGCTTGGCATGTCAGATTTTCTCGGTGGGTATCGGGGTGAGGC 1860
Db 1801 TGTTCGAGATGAGCTTGGCATGTCAGATTTTCTCGGTGGGTATCGGGGTGAGGC 1860
Qy 1861 AGGATCCTTGAACTGCCAAACGCTCTGGAGCTTCTTCCCTTTCTTCCAGCGATCCCGAGC 1920
Db 1861 AGGATCCTTGAACTGCCAAACGCTCTGGAGCTTCTTCCCTTTCTTCCAGCGATCCCGAGC 1920
Qy 1921 GTAAATTTCTCCGGGACACCCCTTTTTCGACAGACAATAAGCCCTTGAGCGCTAG 1980
Db 1921 GTAAATTTCTCCGGGACACCCCTTTTTCGACAGACAATAAGCCCTTTGAGCGCTAG 1980
Qy 1981 CTCTCAAGCTTCTCCGGGGATAGCGCAATGCGCTCCAGAGGGGGAAGTATTCTCTCG 2040
Db 1981 CTCTCAAGCTTCTCCGGGGATAGCGCAATGCGCTCCAGAGGGGGAAGTATTCTCTCG 2040

Qy 2041 CCAAGGCTTCCCGTAGGGCCATCTTTGGTTTCTCCAGGAGCATGCGAGGATTTGGTGT 2100
Db 2041 CCAAGGCTTCCCGTAGGGCCATCTTTGGTTTCTCCAGGAGCATGCGAGGATTTGGTGT 2100
Qy 2101 GTACCGTTCCCGTTCTCGTCTACAAAGGGAAAAAGCTAGCATCTCTCTTCCGAATA 2160
Db 2101 GTACCGTTCCCGTTCTCGTCTACAAAGGGAAAAAGCTAGCATCTCTCTTCCGAATA 2160
Qy 2161 GGGCTAGCGGATTCGTTTCCAAACGTAAGTCCCGCTTTTGGAGTAGACGAGGATCATGTC 2220
Db 2161 GGGCTAGCGGATTCGTTTCCAAACGTAAGTCCCGCTTTTGGAGTAGACGAGGATCATGTC 2220
Qy 2221 CTTTTCGATCCGAGGCTTACCGGAAAAAGTTTTCGGGATTTGAAGCGATGCGGCGAT 2280
Db 2221 CTTTTCGATCCGAGGCTTACCGGAAAAAGTTTTCGGGATTTGAAGCGATGCGGCGAT 2280
Qy 2281 ATGTTTAAACGAAGTTTCGCGGCGCAAGACCTCATCAAGGATGAGCTTCACTTCGAACCC 2340
Db 2281 ATGTTTAAACGAAGTTTCGCGGCGCAAGACCTCATCAAGGATGAGCTTCACTTCGAACCC 2340
Qy 2341 GTATTTCTCGTCTATGTGAACGAAGATCAGTCTGAGTCCGCCATCAGCTCCCTGAGAAG 2400
Db 2341 GTATTTCTCGTCTATGTGAACGAAGATCAGTCTGAGTCCGCCATCAGCTCCCTGAGAAG 2400
Qy 2401 TATCAAGCGCTCCCTCAGGAACCTCCAAACTCAGGACCATCGAGGCTGTCTAGGCC 2460
Db 2401 TATCAAGCGCTCCCTCAGGAACCTCCAAACTCAGGACCATCGAGGCTGTCTAGGCC 2460
Qy 2461 CAATCTGACCGTTTTCGGCTGGCTGACGCTAGCAAGCGCATCTGTTTTCATCGCGCGCAAC 2520
Db 2461 CAATCTGACCGTTTTCGGCTGGCTGACGCTAGCAAGCGCATCTGTTTTCATCGCGCGCAAC 2520
Qy 2521 GAGAAATCTGTCGGCGGTTCCATAAGCGGCTCAATATAGACCACTTGAACCTTCCCGC 2580
Db 2521 GAGAAATCTGTCGGCGGTTCCATAAGCGGCTCAATATAGACCACTTGAACCTTCCCGC 2580
Qy 2581 ATACCCACAGGCTCCCGGAGCATCCACCGGAGAACTGACCGTTTCCCGCAAAAGTA 2640
Db 2581 ATACCCACAGGCTCCCGGAGCATCCACCGGAGAACTGACCGTTTCCCGCAAAAGTA 2640
Qy 2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGCATTTTCCCTTAGGAAGAGGGTTTC 2700
Db 2641 GGTGCCAATAGGATCAATCTCAAAAAGGGGGCATTTTCCCTTAGGAAGAGGGTTTC 2700
Qy 2701 TTTTCGCAAAACGAAGTTTGGGCTGAGTCAAGAACTCTCTCTCATCGGTTTTC 2760
Db 2701 TTTTCGCAAAACGAAGTTTGGGCTGAGTCAAGAACTCTCTCTCATCGGTTTTC 2760
Qy 2761 GGGTAGACCAACCTTAAAGGGCGAAGTTCCGAGGTTTTCGAGGCTTTCAGGGGCTTT 2820
Db 2761 GGGTAGACCAACCTTAAAGGGCGAAGTTCCGAGGTTTTCGAGGCTTTCAGGGGCTTT 2820
Qy 2821 TCGGGTCAAAACCGAGGTAGTCTGCTCATTTTCCTCCCGACAGGCTTCTAAGCAG 2880
Db 2821 TCGGGTCAAAACCGAGGTAGTCTGCTCATTTTTCCTCCCGACAGGCTTCTAAGCAG 2880
Qy 2881 ACCTCATCAACCAACCCCTCAGCACTCCAAACCAAGGAATCCGCAAAAGGCGGCTTACC 2940
Db 2881 ACCTCATCAACCAACCCCTCAGCACTCCAAACCAAGGAATCCGCAAAAGGCGGCTTACC 2940
Qy 2941 TTTTGAAGCGGTATCTTCCCTGATAGACCTTTCGGATCTCTCAGGGTTCACCCGA 3000
Db 2941 TTTTGAAGCGGTATCTTCCCTGATAGACCTTTCGGATCTCTCAGGGTTCACCCGA 3000
Qy 3001 AGGATGCTCAAGCTCCTCGGGGTGAGTACAGGCTTTCATCTCATGACACAACCT 3060
Db 3001 AGGATGCTCAAGCTCCTCGGGGTGAGTACAGGCTTTCATCTCATGACACAACCT 3060
Qy 3061 TACCCCAAGAGGACCAACATGCAATATGCGGAAAGTAGACAACAGAGCAACAAAGCTT 3120
Db 3061 TACCCCAAGAGGACCAACATGCAATATGCGGAAAGTAGACAACAGAGCAACAAAGCTT 3120


```
Db 5281 GTGGGGATAGTCTCAAGTACTCCCAAGGTTCTCGGGCCGTGTGTGGGAGAGAC 5340
Qy 5341 AAAGGGTGGGCAAAAGTTCAFTTTGTACTTTAGGACGATTAATTTAGCACCTGATAA 5400
Db 5341 AAAGGGTGGGCAAAAGTTCAFTTTGTACTTTAGGACGATTAATTTAGCACCTGATAA 5400
Qy 5401 CTTGAGGCGGTTAAGAGGGCTCACTCGGAGCGGTGGAAGGAGACCTGGGCGTG 5460
Db 5401 CTTGAGGCGGTTAAGAGGGCTCACTCGGAGCGGTGGAAGGAGACCTGGGCGTG 5460
Qy 5461 GAAGAAGACGAACCCCGATTTTGGGAAGTCTCCCTCCAGTTTGTATGATGAACGTTGGCA 5520
Db 5461 GAAGAAGACGAACCCCGATTTTGGGAAGTCTCCCTCCAGTTTGTATGATGAACGTTGGCA 5520
Qy 5521 GGAAGCGGCGCAGGATGTCTTCATCGCGGCTCGAACCCTCGACACATATAAACTTTCG 5580
Db 5521 GGAAGCGGCGCAGGATGTCTTCATCGCGGCTCGAACCCTCGACACATATAAACTTTCG 5580
Qy 5581 TGTTCGTAGGCGCAAGAGTGTATGTATGAGGTAACTTTCGGAGTACAAAGTGCCTCAA 5640
Db 5581 TGTTCGTAGGCGCAAGAGTGTATGTATGAGGTAACTTTCGGAGTACAAAGTGCCTCAA 5640
Qy 5641 GCCGCTTTCCCAACGCTCAAACTCTAGGTCAGTGGTTCAGTTTCTGAAAACT 5700
Db 5641 GCCGCTTTCCCAACGCTCAAACTCTAGGTCAGTGGTTCAGTTTCTGAAAACT 5700
Qy 5701 CTAGCTTTTCAGGTCATCTCCTCACCCTCTAGCAGTACTCTGGAAGTAACTTTG 5760
Db 5701 CTAGCTTTTCAGGTCATCTCCTCACCCTCTAGCAGTACTCTGGAAGTAACTTTG 5760
Qy 5761 ACACAGCGGCAAGTCTAGGTCCTCCAGTCAGTTGTCCTGGACGCGTGAGAGGGGA 5820
Db 5761 ACACAGCGGCAAGTCTAGGTCCTCCAGTCAGTTGTCCTGGACGCGTGAGAGGGGA 5820
Qy 5821 GGGGCTTGTGTAGAGACGAGAGACCC 5849
Db 5821 GGGGCTTGTGTAGAGACGAGAGACCC 5849

RESULT 3
U89376/c
LOCUS U89376 1600 bp DNA linear BCT 06-OCT-1997
DEFINITION Thermus sp. Y845 replication protein (repT) gene, complete cds.
ACCESSION U89376
VERSION U89376.1 GI:2465516
KEYWORDS
SOURCE Thermus sp. Y845
ORGANISM Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae; Thermus.
REFERENCE
1 (bases 1 to 1600)
AUTHORS Wayne, J. and Xu, S.-Y.
TITLE Identification of a thermophilic plasmid origin and its cloning within a new Thermus-E. coli shuttle vector
JOURNAL Gene 195 (2), 321-328 (1997)
MEDLINE 97449309
PUBMED 9305778
REFERENCE
2 (bases 1 to 1600)
AUTHORS Wayne, J. and Xu, S.-Y.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1997) Research, New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA
FEATURES
source
1. .1600
/organism="Thermus sp. Y845"
/mol_type="genomic DNA"
/strain="Y845"
/db_xref="taxon:64227"
220..225
/notes="putative"
-35_signal
243..248
-10_signal
278..283
-35_signal
```

```
-10_signal
gene
RBS
CDS
/note="putative"
302..307
/note="putative"
332..1369
/gene="repT"
332..336
/gene="repT"
344..1369
/gene="repT"
/functions="required for plasmid replication"
/note="repT"
/codon_start=1
/transl_table=11
/product="replication protein"
/protein_id="AA872144.1"
/db_xref="GI:2465517"
/translation="MKNEKTFPEELYEAEETHNDTATRGSDRGSEDFFLANDPPDP
GGANRLAKGFTYKEALRIALPEKHOEAFLSVGGAPPVPAEPFVGVNVCQAVQDPQ
KLLELQEIARSTPYGNRELKRVGVVFMVPLEMLNLGLVTRCTHAWKVLKFK
GLVATDVLHQTNGERRALGTJWVRLRFGKRLTDDYIYWRNLALDMANGVLSFN
WVAYODHGIRPTLDVLVLAOGKRVMPNTKTVAVDLGLILVPEVERSKLPLILILI
ATYIALDLDLRRSRFRYAGLLWAVARGELPAQVLFVAVLMRVIRDYTDGHLTRPGALV
KTLKEAS"
1158..1166
/gene="repT"
/note="putative; encodes DnaA box"
1293..1301
/gene="repT"
/note="putative; encodes DnaA box"

ORIGIN
Query Match 27.4%; Score 1600; DB 1; Length 1600;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1600; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2901 CAGCACTCCAAACCAAGGATCCGCAAGGCGGCTACTTTTGAGCCGATCTTCCC 2960
Db 1600 CAGCACTCCAAACCAAGGATCCGCAAGGCGGCTACTTTTGAGCCGATCTTCCC 1541
Qy 2961 CTGACGTATAGACTTCGGATCGTCTCAGGGTGCACCCGGAAGGATGTCTGCAAGCTCCTC 3020
Db 1540 CTGACGTATAGACTTCGGATCGTCTCAGGGTGCACCCGGAAGGATGTCTGCAAGCTCCTC 1481
Qy 3021 GGGGTGAGTACACGGGCTTCATCTCATGACACACCTTACCCACAGAGAGACACAC 3080
Db 1480 GGGGTGAGTACACGGGCTTCATCTCATGACACACCTTACCCACAGAGAGACACAC 1421
Qy 3081 ATGCAACTATGGGCAAGTAGACAAACAGACACCAAAAGCTTGGGCCACTCTCTCAGGAGGC 3140
Db 1420 ATGCAACTATGGGCAAGTAGACAAACAGACACCAAAAGCTTGGGCCACTCTCTCAGGAGGC 1361
Qy 3141 CTCCTTGAGGCTTTCATAGGTACGCTCCGGTCTGTGTGATGCGCATCCGTGTAATC 3200
Db 1360 CTCCTTGAGGCTTTCATAGGTACGCTCCGGTCTGTGTGATGCGCATCCGTGTAATC 1301
Qy 3201 TCGGTAACCCGATTAGGACGCAATAGATATTCGGGGGAGTTTCACTCCCTGGCCAC 3260
Db 1300 TCGGTAACCCGATTAGGACGCAATAGATATTCGGGGGAGTTTCACTCCCTGGCCAC 1241
Qy 3261 AGCCACAGCAAGCTGCATAGAAACGTTTGAACACGCTCATCTAGAGATCGGCAAT 3320
Db 1240 AGCCACAGCAAGCTGCATAGAAACGTTTGAACACGCTCATCTAGAGATCGGCAAT 1181
Qy 3321 GTACGTAGCAATGAGGTGATAGGGCCGGAGTTTGGACCGCTCCACCTCGGGGAGGAC 3380
Db 1180 GTACGTAGCAATGAGGTGATAGGGCCGGAGTTTGGACCGCTCCACCTCGGGGAGGAC 1121
Qy 3381 CAGGATGAGGCCCGAGGTCAACGGCCACGCTTCTTGGTGTGGGCATCACCTTTTCCCTCTG 3440
Db 1120 CAGGATGAGGCCCGAGGTCAACGGCCACGCTTCTTGGTGTGGGCATCACCTTTTCCCTCTG 1061
Qy 3441 AGCCGAGGACGAGACGCTCCAGGTTGGGGCGGATTCGTTGGTCTGTAGGCTTGAC 3500
Db 3500 AGCCGAGGACGAGACGCTCCAGGTTGGGGCGGATTCGTTGGTCTGTAGGCTTGAC 3500
```


Db 1060 AGCCAGGAGGACAGCAGCTCCAGGGTGGGGCGATTCCCGTGTCTGTAGGCTTGAC 1001
 Qy 3501 CCAATTGAAGAGAGACAGCGCTTGGCCATGTCTAGGGCGAGAGTTCCTCCAGGGGTAGAT 3560
 Db 1000 CCAATTGAAGAGAGACAGCGCTTGGCCATGTCTAGGGCGAGAGTTCCTCCAGGGGTAGAT 941
 Qy 3561 GTAGTCGTCCAGGGTGAAGCTGCTTCCCTTGGCTCAGCGCGAGCGCCCAAAAGGGTGCC 3620
 Db 940 GTAGTCGTCCAGGGTGAAGCTGCTTCCCTTGGCTCAGCGCGAGCGCCCAAAAGGGTGCC 881
 Qy 3621 GATGCCCGGCGCTCCCGCTTGAAGCTGCTTGGTGAAGAGCTGCTGAGCGCCCAAGAGCGCTT 3680
 Db 880 GATGCCCGGCGCTCCCGCTTGAAGCTGCTTGGTGAAGAGCTGCTGAGCGCCCAAGAGCGCTT 821
 Qy 3681 TTTCTCAAGGACCTTCTCCAGGCGTGAAGCTGCTGCGGGTGACCCCGAGGTTGAGGC 3740
 Db 820 TTTCTCAAGGACCTTCTCCAGGCGTGAAGCTGCTGCGGGTGACCCCGAGGTTGAGGC 761
 Qy 3741 CAACATCTCCAGGGGACCATGAAGAGACCGCTCCCGCTTCCCTCCAGAGCTCCCGGTT 3800
 Db 760 CAACATCTCCAGGGGACCATGAAGAGACCGCTCCCGCTTCCCTCCAGAGCTCCCGGTT 701
 Qy 3801 CCGCTAGGGATGTGGAGCGGGCAATCTCTGAGAGATTCAGAGAGCTTCTGAGGCC 3860
 Db 700 CCGCTAGGGATGTGGAGCGGGCAATCTCTGAGAGATTCAGAGAGCTTCTGAGGCC 641
 Qy 3861 GTCTGGAGCGCTTGACATATCCCAACCGGGGCTTCACTGGTGTATAGGGGGGC 3920
 Db 640 GTCTGGAGCGCTTGACATATCCCAACCGGGGCTTCACTGGTGTATAGGGGGGC 581
 Qy 3921 CCCAACAGAGAAAGGAGCTCATGCTCTTCTCGGTAAAGCAATCTTAAGTGCTC 3980
 Db 580 CCCAACAGAGAAAGGAGCTCATGCTCTTCTCGGTAAAGCAATCTTAAGTGCTC 521
 Qy 3981 TTTTGTGTATGAAGCGCTTGGAGCGGCAATCTTCTCGGTAAAGCAATCTTAAGTGCTC 4040
 Db 520 TTTTGTGTATGAAGCGCTTGGAGCGGCAATCTTCTCGGTAAAGCAATCTTAAGTGCTC 461
 Qy 4041 GGTGCCAAGAGAGTCTCTGACCCCTATCTGACCCCTAGTGGCATCGGTGTGTC 4100
 Db 460 GGTGCCAAGAGAGTCTCTGACCCCTATCTGACCCCTAGTGGCATCGGTGTGTC 401
 Qy 4101 GTGGGTTTCTTAAGCGCTGTAAAGCTCTTCAAGAGGTTTCTTCTTCAAGCT 4160
 Db 400 GTGGGTTTCTTAAGCGCTGTAAAGCTCTTCAAGAGGTTTCTTCTTCAAGCT 341
 Qy 4161 CGGACCTCTTGTATCTGAGCGCGGCGTATACCTAGTCTTGGGGGTGATCCGGGG 4220
 Db 340 CGGACCTCTTGTATCTGAGCGCGGCGTATACCTAGTCTTGGGGGTGATCCGGGG 281
 Qy 4221 CAACCGCTCGGTTTGGCTTTTATGGGTCCAAATAAACCGTCAGCCAGCGCTGGCA 4280
 Db 280 CAACCGCTCGGTTTGGCTTTTATGGGTCCAAATAAACCGTCAGCCAGCGCTGGCA 221
 Qy 4281 ATCCCGCTCTTAAGGCGGTTATAGCGCTGCTAGGAGGGGTAGTACTTCTTACC 4340
 Db 220 ATCCCGCTCTTAAGGCGGTTATAGCGCTGCTAGGAGGGGTAGTACTTCTTACC 161
 Qy 4341 CCCCTAGGCTTGGAGAGCTTATAGAGTCTCTAGGCGCTCTGCTGGGGGTGAGGGGTAA 4400
 Db 160 CCCCTAGGCTTGGAGAGCTTATAGAGTCTCTAGGCGCTCTGCTGGGGGTGAGGGGTAA 101
 Qy 4401 CCTCATGCGAGCGCGGCTCGGAGCTCTGGAGGAGGCTCCATAGCTTCTGCTGG 4460
 Db 100 CCTCATGCGAGCGCGGCTCGGAGCTCTGGAGGAGGCTCCATAGCTTCTGCTGG 41
 Qy 4461 TGGAGGTTTGAAGGGGTTCACTAATCATACGCTAGC 4500
 Db 40 TGGAGGTTTGAAGGGGTTCACTAATCATACGCTAGC 1

RESULT 4
 AR031809/c

LOCUS AR031809 1242 bp DNA linear PAT 29-SEP-1999
 DEFINITION Sequence 1 from patent US 5864422.
 ACCESSION AR031809
 VERSION AR031809.1 GI:5946098
 KEYWORDS Unknown.
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1242)
 AUTHORS Wayne, J. and Xu, S.-Y.
 TITLE Method for cloning and producing the Tsp45I restriction
 endonuclease in *E. coli*
 JOURNAL Patent: US 5864422-A 1 02-FEB-1999;
 FEATURES
 Location/Qualifiers
 source 1..1242
 /organism="unknown"
 /mol_type="unassigned DNA"
 ORIGIN
 Query Match 21.2%; Score 1242; DB 6; Length 1242;
 Best Local Similarity 100.0%; Pred. No. 4.3e-256;
 Matches 1242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1609 CTAGAAGGCGGACACATCTCAAACTTGTGCTGTAGCTGGGGAATCTCTAAACACCT 1668
 Db 1242 CTAGAAGGCGGACACATCTCAAACTTGTGCTGTAGCTGGGGAATCTCTAAACACCT 1183
 Qy 1669 TCTAGTGAAGGCTTTGACCGCTCCAGGAGCATCTATGCCAGATGGATCGCGCTTAA 1728
 Db 1182 TCTAGTGAAGGCTTTGACCGCTCCAGGAGCATCTATGCCAGATGGATCGCGCTTAA 1123
 Qy 1729 GAGGGTGAAGGCTATAGGCTAGTACCGGAGCTCGGAAGGATCGAGCATTAATCCCC 1788
 Db 1122 GAGGGTGAAGGCTATAGGCTAGTACCGGAGCTCGGAAGGATCGAGCATTAATCCCC 1063
 Qy 1789 CTCGTACTCCCTGTTTGGACGATGAGCTTGCAGATGTCAGATTTTCTCGTGGGTA 1848
 Db 1062 CTCGTACTCCCTGTTTGGACGATGAGCTTGCAGATGTCAGATTTTCTCGTGGGTA 1003
 Qy 1849 TCGCGGTACGGAGGATCTTGAACCTGCCAAACGCTCTGAGCTTCTTCCCTTCTTCA 1908
 Db 1002 TCGCGGTACGGAGGATCTTGAACCTGCCAAACGCTCTGAGCTTCTTCCCTTCTTCA 943
 Qy 1909 GCGATCCCGAGGTAAACTTCTTCGCGGACCGCTTCTTTCGACGACACATAGCCC 1968
 Db 942 GCGATCCCGAGGTAAACTTCTTCGCGGACCGCTTCTTTCGACGACACATAGCCC 883
 Qy 1969 TTGAGGCTCTAGCTCGTCAAGCTTCTTCGCGGAGATAGCGCAATGCCGTCAGGGGG 2028
 Db 882 TTGAGGCTCTAGCTCGTCAAGCTTCTTCGCGGAGATAGCGCAATGCCGTCAGGGGG 823
 Qy 2029 AGTATTCCTCGCAAGGCTTCCGCTAGGCGCATCTTGGTTCCTCAGGAGCATGCG 2088
 Db 822 AGTATTCCTCGCAAGGCTTCCGCTAGGCGCATCTTGGTTCCTCAGGAGCATGCG 763
 Qy 2089 GGGATTGGTGTACCGTTCCCGTTCGCTCTCAAAAGGGAAGCTTAGCGATCTC 2148
 Db 762 GGGATTGGTGTACCGTTCCCGTTCGCTCTCAAAAGGGAAGCTTAGCGATCTC 703
 Qy 2149 CTCTCCGAATAGGGCTAGCGCATCTTCAAAAGCTAGTCCCGCTTTTGGAGTAGAC 2208
 Db 702 CTCTCCGAATAGGGCTAGCGCATCTTCAAAAGCTAGTCCCGCTTTTGGAGTAGAC 643
 Qy 2209 GAGGATCATGTCTTTTTCGATCCGAGGCTTACGGGAAGTCTTTCGGATTTCGAGC 2268
 Db 642 GAGGATCATGTCTTTTTCGATCCGAGGCTTACGGGAAGTCTTTCGGATTTCGAGC 593
 Qy 2269 GATCGGGCGATATGTTTAAACGAAGTTTCGCGGCAAGAGCTCATCAAGGATGAGCTT 2328
 Db 582 GATCGGGCGATATGTTTAAACGAAGTTTCGCGGCAAGAGCTCATCAAGGATGAGCTT 523
 Qy 2329 CACCTCGAACCGGCTATTTCTGCTATGTGAACGAGATCAGTCTCGGATCGCCATCAG 2388

```

Db      522  CACCTCGAACCCGCTATTTCTGCTGTATGTAACGAAGATCAGTCCCTGAGTCCGCGCATCAG 463
QY      2389  CTCCTCGAGAAGTATCAAGCGCTCCCTCAGGAATCTCACAACCTGAGGACCAATCAGAGGT 2448
Db      462  CTCCTCGAGAAGTATCAGCGCTCCCTCAGGAATCTCACAACCTGAGGACCAATCAGAGGT 403
QY      2449  GTCATGTAGCCCAACTGACCGCTTTTGGCTGGCTGACGGTAGCAACCGGATCTGTTTC 2508
Db      402  GTCATGTAGCCCAACTGACCGCTTTTGGCTGGCTGACGGTAGCAACCGGATCTGTTTC 343
QY      2509  ATCGCGCCCAACGAGAACTGCTGGCGGTTCATTAAGCGGCTCAATATAGCAAACTG 2568
Db      342  ATCGCGCCCAACGAGAACTGCTGGCGGTTCATTAAGCGGCTCAATATAGCAAACTG 283
QY      2569  GACCTTCCCGCATACCCACAGGCTCCCGGAGCATCCACCGAGAACCTGACCGTTTTC 2628
Db      282  GACCTTCCCGCATACCCACAGGCTCCCGGAGCATCCACCGAGAACCTGACCGTTTTC 223
QY      2629  CCCCAAAAGTAGTCCCAATAGGATCAATCTCAAAAGGGGGCAATTTCCCGCTAGGAA 2688
Db      222  CCCCAAAAGTAGTCCCAATAGGATCAATCTCAAAAGGGGGCAATTTCCCGCTAGGAA 163
QY      2689  GAGGAGGGTTTCTTTTCGCAAAACAAGTTGTGGGTGGCTGATCAAGAATCTCCTTCTC 2748
Db      162  GAGGAGGGTTTCTTTTCGCAAAACAAGTTGTGGGTGGCTGATCAAGAATCTCCTTCTC 103
QY      2749  ATCGCGTTTTCGGGTAGACCAACTAAGGGCGAGGTTCCGAGTTTTCGAGGCTTT 2808
Db      102  ATCGCGTTTTCGGGTAGACCAACTAAGGGCGAGGTTCCGAGTTTTCGAGGCTTT 43
QY      2809  CAAGGGGGCTTTTCGGGTCAAAACAGGCTAGCTACGGCTCAT 2850
Db      42  CAAGGGGGCTTTTCGGGTCAAAACAGGCTAGCTACGGCTCAT 1

RESULT 5
AR139424/c
LOCUS      AR139424      1026 bp      DNA      linear      PAT 16-JUN-2001
DEFINITION Sequence 4 from patent US 6207377.
ACCESSION AR139424
VERSION AR139424.1 GI:14481920
KEYWORDS   Location/Qualifiers
SOURCE     Unknown.
           Unclassified.
REFERENCE  1 (bases 1 to 1026)
AUTHORS   Wayne, J. and Xu, S.-Y.
TITLE     Method for construction of thermus-E. coli shuttle vectors and
           identification of two Thermus plasmid replication origins
JOURNAL   Patent: US 6207377-A 4 27-MAR-2001;
FEATURES   Location/Qualifiers
           source      1..1026
                       /organism="unknown"
                       /mol_type="unassigned DNA"
ORIGIN
Query Match      17.5%; Score 1026; DB 6; Length 1026;
Best Local Similarity 100.0%; Pred. No. 1.2e-209;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3132  TCAGGAGGCTCTTCAGGCTTCTACTAGGTACGCTCCCGGTGTCAGATGCCATC 3191
Db      1026  TCAGGAGGCTCTTCAGGCTTCTACTAGGTACGCTCCCGGTGTCAGATGCCATC 967
QY      3192  CGTGTAACTCTCGGATAACCGCATTAGGACGGCAATAGATATTCGGGGAGTTACCC 3251
Db      966  CGTGTAACTCTCGGATAACCGCATTAGGACGGCAATAGATATTCGGGGAGTTACCC 907
QY      3252  CTTGGCCACAGCCCAAGCAGGCTGCATAGAAAGCTTTTGAACGCGTCTAGGAG 3311
Db      906  CTTGGCCACAGCCCAAGCAGGCTGCATAGAAAGCTTTTGAACGCGTCTAGGAG 847
QY      3312  ATCGGCAATGTACGTAGCAATAGGGTGATAAGGGCGGGAGTTTGAACGCTCCACCTC 3371

```

```

Db      846  ATCGECAATGTACGTAGCAATGAGGGTGATAAGGGCGGGAGTTTGAACGCTCCACCTC 787
QY      3372  GGGGAGGACCAAGGATGAGGCCAGGTCACCGGCAAGGTTTGGTGTGGGATCACCTC 3431
Db      786  GGGGAGGACCAAGGATGAGGCCAGGTCACCGGCAAGGTTTGGTGTGGGATCACCTC 727
QY      3432  TTTCCCTGAGCCAGAGGACAGACAGTCCAGGGTGGGCGGATTCCTGGTCTGGTA 3491
Db      726  TTTCCCTGAGCCAGAGGACAGACAGTCCAGGGTGGGCGGATTCCTGGTCTGGTA 667
QY      3492  GGCCTTGACCCAGTTGAAGGAGAGACGCGCTTGGCCATGCTAGGGCGAGGTTCTCCA 3551
Db      666  GGCCTTGACCCAGTTGAAGGAGAGACGCGCTTGGCCATGCTAGGGCGAGGTTCTCCA 607
QY      3552  GGGGTAGATGTAGTCTCCAGGGTGAAGCTTTCCTTGGCTCAGCGGACGGCCCA 3611
Db      606  GGGGTAGATGTAGTCTCCAGGGTGAAGCTTTCCTTGGCTCAGCGGACGGCCCA 547
QY      3612  AAGGTCGCCATGCGCCGCGCTCCCGTTGACGTTTGGTGAAGGACGTCGCTGGCCAC 3671
Db      546  AAGGTCGCCATGCGCCGCGCTCCCGTTGACGTTTGGTGAAGGACGTCGCTGGCCAC 487
QY      3672  CAGGCCCTTTTCTCAAGGACCTTCTTCCAGGGTGAAGCTTTCGGGTGACCCCGAG 3731
Db      486  CAGGCCCTTTTCTCAAGGACCTTCTTCCAGGGTGAAGCTTTCGGGTGACCCCGAG 427
QY      3732  GTTGAGGGCCCAACATCTCCAGGGGACCATGAAGACGCGTCCCGCTTCTCCAGAG 3791
Db      426  GTTGAGGGCCCAACATCTCCAGGGGACCATGAAGACGCGTCCCGCTTCTCCAGAG 367
QY      3792  CTCGCCGTTCGCGTAGGGGATGTTGGAGCGGGCAATCTCTCGAGGAGTTTCCAGAACTT 3851
Db      366  CTCGCCGTTCGCGTAGGGGATGTTGGAGCGGGCAATCTCTCGAGGAGTTTCCAGAACTT 307
QY      3852  CTGAGGCCCTCTCTGACCGCTTGACATATCTCCAAACGGGGGTTTACGCTGGTGTAT 3911
Db      306  CTGAGGCCCTCTCTGACCGCTTGACATATCTCCAAACGGGGGTTTACGCTGGTGTAT 247
QY      3912  AGGGGGGGCCCCAACAGAGGAAAGGAAAGCTCATGCTCTTCTCGGGTAAAGCAATCCT 3971
Db      246  AGGGGGGGCCCCAACAGAGGAAAGGAAAGCTCATGCTCTTCTCGGGTAAAGCAATCCT 187
QY      3972  AAGTGCCTCTTTTGGTATCTAAAGCCCTTCGAGCGGATTTTCGGCACTCATCTGG 4031
Db      186  AAGTGCCTCTTTTGGTATCTAAAGCCCTTCGAGCGGATTTTCGGCACTCATCTGG 127
QY      4032  AGGGGGTTCGGTGGCCAGAGGAGTCTCTGACCCCTCTGACCCCTCTGACCCCTAGTGCATC 4091
Db      126  AGGGGGTTCGGTGGCCAGAGGAGTCTCTGACCCCTCTGACCCCTCTGACCCCTAGTGCATC 67
QY      4092  GGTGTTCTCGTGGGTTTCTCTTAAAGCTCTGTAAGCTCTTCAAGAGGTTTTCGTT 4151
Db      66  GGTGTTCTCGTGGGTTTCTCTTAAAGCTCTGTAAGCTCTTCAAGAGGTTTTCGTT 7
QY      4152  CTTTAC 4157
Db      6  CTTTAC 1

RESULT 6
AR031810
LOCUS      AR031810      999 bp      DNA      linear      PAT 29-SEP-1999
DEFINITION Sequence 3 from patent US 5866422.
ACCESSION AR031810
VERSION AR031810.1 GI:5946099
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 999)
AUTHORS   Wayne, J. and Xu, S.-Y.
TITLE     Method for cloning and producing the Tsp45I restriction

```


engonuclease in E. coli	
JOURNAL	Patent: US 596422-A 3 02-FEB-1999;
FEATURES	Location/Qualifiers
source	1..999
	/organism="unknown"
	/mol_type="unassigned DNA"
ORIGIN	
Query Match	17.1%; Score 999; DB 6; Length 999;
Best Local Similarity	100.0%; Pred. No. 7.6e-204;
Matches 999; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	614 ATGCAACAGATGGCGGATGGAAGCTGTGGACACAGAGAGCGTGTGAGCTTCTGGAGAAG 673
Db	1 ATGCAACAGATGGCGGATGGAAGCTGTGGACACAGAGAGCGTGTGAGCTTCTGGAGAAG 60
QY	674 GGGTATTGGTAACTACTCTACGCTCTATAAAGGGGAAAGTGGCTCTTCGAGTCACTA 733
Db	61 GGGTATTGGTAACTACTCTACGCTCTATAAAGGGGAAAGTGGCTCTTCGAGTCACTA 120
QY	734 CCAGAGGAGGTAGAGGAAAAAATTTCGCGAGGCTTACAGGCATACGAGGGGAGGACGAGAT 793
Db	121 CCAGAGGAGGTAGAGGAAAAAATTTCGCGAGGCTTACAGGCATACGAGGGGAGGACGAGAT 180
QY	794 AGTCGGAGGAGGAAACGAAACTGTGTGAGCCGTGCTAAATGCCAGAAAAAGGTCGAG 853
Db	181 AGTCGGAGGAGGAAACGAAACTGTGTGAGCCGTGCTAAATGCCAGAAAAAGGTCGAG 240
QY	854 CGGTCCCTCTCAATCACCCCTACTCGCTGTGCTTGTCTACTACTCTGTTTCGAAAAAGCA 913
Db	241 CGGTCCCTCTCAATCACCCCTACTCGCTGTGCTTGTCTACTACTCTGTTTCGAAAAAGCA 300
QY	914 GAAAAAGCGAACAGGCCCTTTAGAGGAGCATGTGAGGAGGTGTGCTTAAGACCCACGAA 973
Db	301 GAAAAAGCGAACAGGCCCTTTAGAGGAGCATGTGAGGAGGTGTGCTTAAGACCCACGAA 360
QY	974 ACCATCCGCTCTCGGCAAGGAGCGCAAGAGAGGCGCTAGAGCGCTTCATCCAAAGG 1033
Db	361 ACCATCCGCTCTCGGCAAGGAGCGCAAGAGAGGCGCTAGAGCGCTTCATCCAAAGG 420
QY	1034 CTCAGGAGCGCTCCGGAATAAATACCGCAGATAGGCGCGATGTTCAAAAGGTGTACAA 1093
Db	421 CTCAGGAGCGCTCCGGAATAAATACCGCAGATAGGCGCGATGTTCAAAAGGTGTACAA 480
QY	1094 GAAGAGCTAAAGGGGAAAAATAGAGAGAGGCTTCCAGGCCCTTACCAACCAAGATTTGT 1153
Db	481 GAAGAGCTAAAGGGGAAAAATAGAGAGAGGCTTCCAGGCCCTTACCAACCAAGATTTGT 540
QY	1154 GTAGTATCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTTCGGAGAGAGAGCG 1213
Db	541 GTAGTATCCCTGAAAAAAGTAAACCGGAGCAAGCACCCCTTATTTCGGAGAGAGAGCG 600
QY	1214 GGCATCATATATACACGGGATCGGATGAAGCTTTGAAAGATGCGCGCAAGGAAAAACCTG 1273
Db	601 GGCATCATATATACACGGGATCGGATGAAGCTTTGAAAGATGCGCGCAAGGAAAAACCTG 660
QY	1274 GGCCTTGGGAGAGAGAGAACTAGGACCAAGGCGTAGATTTCTAGCTGGTCACTCGG 1333
Db	661 GGCCTTGGGAGAGAGAGAACTAGGACCAAGGCGTAGATTTCTAGCTGGTCACTCGG 720
QY	1334 CGTAGCCCTTGAGAGACATGGCACTTAACAGGAGAGTCAAGTTTCAATCCGACTTTGGC 1393
Db	721 CGTAGCCCTTGAGAGACATGGCACTTAACAGGAGAGTCAAGTTTCAATCCGACTTTGGC 780
QY	1394 GGAACCAAGAGACACAGAACTAGTAGCAAGGCTTCCATAAGGTTGGACCTTGAGAG 1453
Db	781 GGAACCAAGAGACACAGAACTAGTAGCAAGGCTTCCATAAGGTTGGACCTTGAGAG 840
QY	1454 AGGCACATAGGAATAGTGTGTGGACGGAATCCCTGTGAGCAAGTGTTCGTGGGTGG 1513
Db	841 AGGCACATAGGAATAGTGTGTGGACGGAATCCCTGTGAGCAAGTGTTCGTGGGTGG 900
QY	1514 GCCGAGCTGGGGAAGAAACGATCGTTATCCGCTACTCTCTCCAGACCTGATAGCG 1573
Db	

Db	901 GCCGAGCTGGGAAAGAAACGATCGTTACATCGTACTCTCTTCCAGACCTGATAGCG 960
QY	1574 GAGCTCTACCAAAAGGCTGAAGAAGCCCTGCGCCCTCTAG 1612
Db	961 GAGCTCTACCAAAAGGCTGAAGAAGCCCTGCGCCCTCTAG 999

RESULT 7

AF328924

LOCUS

DEFINITION

Helicobacter pylori strain CR38 cytosine-specific DNA linear BCT 12-NOV-2002

methyitransferase, cytosine-specific DNA methyltransferase, putative HP0052-like protein, type II DNA modification enzyme, and type II restriction enzyme genes, complete cds.

AF328924

AF328924.1 GI:17225546

VERSION

KEYWORDS

SOURCE

ORGANISM

Helicobacter pylori

Helicobacter pylori

Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales; Helicobacteraceae; Helicobacter.

REFERENCE

1 (bases 1 to 4938)

Chanto,G., Occhialini,A., Gras,N., Alm,R.A., Megraud,F. and Marais,A.

AUTHORS

TITLE

Identification of strain-specific genes located outside the plasticity zone in nine clinical isolates of Helicobacter pylori

JOURNAL

PUBMED

12427957

REFERENCE

2 (bases 1 to 4938)

Chanto,G., Occhialini,A., Gras,N., Megraud,F. and Marais,A.

AUTHORS

TITLE

Direct Submission

JOURNAL

Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France

FEATURES

Location/Qualifiers

1..4938

/organism="Helicobacter pylori"

/mol_type="genomic DNA"

/strain="CR38"

/specific_host="Homo sapiens"

/db_xref="taxon:210"

/map="region F; between JHP43/HP50 and JHP47/HP55"

/notes="isolated from patient with chronic gastritis"

complement(1..705)

/notes="HP0051-like protein"

/codon_start=1

/transl_table=11

/product="cytosine-specific DNA methyltransferase"

/protein_id="AAL37468.1"

/db_xref="GI:17225547"

/translation="MKNLSICTKGYFLBEIKERLNAIGYQLSGYQILNAKDYGVQNRE RAPIVCASRSPDFNLPSQSVQDAISDLAYLCSNEGAFESDYLNPIQSSYQALM RKDPSKLYNQATNHSOALEKLIKNKEKCELPKNLHGKQPKSTWGRNLWNKIS PIIDTFDPSNGTNSHPHLRSITPREARIQSPSDNIIFYGKNTSVCKQIGNAVPP LIALAUGKAILSLRK"

complement(750..1073)

/notes="HP0051-like protein"

/codon_start=1

/transl_table=11

/product="cytosine-specific DNA methyltransferase"

/protein_id="AAL37469.1"

/db_xref="GI:17225548"

/translation="MNYKILDLFCGAGGFSAGLEYLKEFDALIGLDCDQKALITFENN HKNAIGVGDITQTEIKEVIELAKLEINMIIGPPCQGFYNGKNGLGLKDPNPLF LEVNI"

complement(1070..2305)

/codon_start=1

/transl_table=11

/product="putative HP0052-like protein"

/protein_id="AAL37470.1"

/db_xref="GI:17225549"

/translations="MLFVSRFLNDINIDPNLGVLLSRFQIKNDRIYGVGSYKSKFTPL GYESKAQVLNALNTLTSTHPIMQSNQSVTKIKGTFTVFILENDLHLDENAFYKLLNL

IIIDNDFNRSLNMLTLMQKRFSLGFFESRGSIDTQRNFLTLDYFFHSPLEPNFKPHY
LIDFPHIPSEALNFRNFIQHEVAQINORNVQFRIYLNWYCHIGLNFYKARIAEH
VFYTLIDYGVYKLVSPPTTKYHNGFIERAHFYKXVYQODLDKRIKESLRBOLG
WIKSEPRKDSKILISFVRLSPNVACDDYDIDKERSFLSLPLVQLTORSDSYTE
LHVVISLGRNLDVLENLCTPCHRALKKGASAEFGKRLIGKILKXNKNLEFVQL
RFETDPLTIDIKIYESLK
complement (2663, .3694)
/function="methyltransferase"
/notes="JHP45-like protein"
/codon_start=1
/transl_table=1
/product="type II DNA modification enzyme"
/protein_id="AAL37471.1"
/db_xref="GI:17225550"
/translation="MQNLIQAEINAIALLFLILNDKMLKGKIDLIYIDPPFATNNHFTI
YKVMLEIDFGIQNFINEITRIKCNPKNFKMGYNIKDMILPYSGKNPIFNEPKIP
YTPDLKSPPKIDKXRYTTPVHAPGEVSGCSKAFKGMPLPKGRHWRTDIATL
ERWDKGLIEYSNNNPKIKIYALEQVGRVQDINEFPDQPYSPYPTERKAGLLDII
KTSNKSIVLDFCGSGGTTLSAFLQLQRKFIQIDNSLAIQACNKLKLETTIKDLFVS
QNPYDFLVF"

CDS

complement (3704, .4489)
/notes="JHP46-like protein"
/codon_start=1
/transl_table=1
/product="type II restriction enzyme"
/protein_id="AAL37472.1"
/db_xref="GI:17225551"
/translation="MNYNTKLISIEYANQSYLDLDFQVYPTIPEGLREIDSKIWSNV
EYHFKKDLALITELLNDLDFPKDSYVAILKROKSALEARNPTINICGRLYGMGL
NKTIFKSPKPTNROIPGMFDWNNKSLGVPEVDLNNFIANDAILRASDNMAE
FASHLNHYHHKGLDVPANPKYIIGEAFLTDGGHONAOFNDAISTIERAPNKAI
KVALDGVLYIBSNKORLDDTITRYNIMSLVLRFLYQI"

CDS

complement (3704, .4489)
/notes="JHP46-like protein"
/codon_start=1
/transl_table=1
/product="type II restriction enzyme"
/protein_id="AAL37472.1"
/db_xref="GI:17225551"
/translation="MNYNTKLISIEYANQSYLDLDFQVYPTIPEGLREIDSKIWSNV
EYHFKKDLALITELLNDLDFPKDSYVAILKROKSALEARNPTINICGRLYGMGL
NKTIFKSPKPTNROIPGMFDWNNKSLGVPEVDLNNFIANDAILRASDNMAE
FASHLNHYHHKGLDVPANPKYIIGEAFLTDGGHONAOFNDAISTIERAPNKAI
KVALDGVLYIBSNKORLDDTITRYNIMSLVLRFLYQI"

ORIGIN

Query Match 3.2%; Score 187.2; DB 1; Length 4938;

Best Local Similarity 52.1%; Pred. No. 2.5e-29;

Matches 476; Conservative 0; Mismatches 423; Indels 15; Gaps 2;

QY	1664	ACCTTCTAGTGAAGGCTTTCACCGCTCCAGGAGGATCTATCGCGATGATGCCGC	1723
DB	2724	AGTTGTGTTTGAAGCTTGGATAGTAATGGAATTATCAATGCTATTAATTTTGT	2783
QY	1724	TTTAAGAGGGGTGAGGCTATAAGCTAGTACCGGAGCTCGAAGGATCGAGCACTAAA	1783
DB	2784	TGCAATAAACAACGATTTTAAAGTTGTTCAGACCAACAACAATCTAAACAATA	2843
QY	1784	TCCCTCTGTTTACCTGTTTGGAGCATGAGCTTGAGCATGTCACAGATTTTTCGGTG	1843
DB	2844	CTATCTTTTATAGAGAGGTTTAAATGATTAAGTCTAATTAATTGAGCGTTTTTCTGTA	2903
QY	1844	GGGTATCGGGGTACGAGGATCTTGAACCTGCAACGCTCTGGAGCTTCTCCCTTC	1903
DB	2904	GGGTAGCTTGATATTGTGGGCTTTTAAATCCAGATGCTTGACTTTTGGC----	2959
QY	1904	TTCAGCGATCCGAGCGTAACTTTCTTCGCGGACCCCGCTTTGACAGCAATA	1963
DB	2960	--AACTTGTCTAAGGCATAAATTTTCTTAGGGTTATTATTGTTAGATTAATCAATC	3017
QY	1964	AGCCCTGAGCGTCTAGCTCGTCAAGCTTCTCGGGGGATAGCGCAATGCCGTCAGGA	2023
DB	3018	AAACCTCTTTTGCCCAAGCTCAAGTGGGCAATATCATAGTGGCCCAATGCCGCTTTT	3077
QY	2024	GGGGAGGTATTCCTCGCCAAAGCCCTTCGGTAGGGCCATCTCTGTTTCTCCAGGAGCA	2083
DB	3078	GGAGGTAGCATACCTTTAAATGCTTTAGAACATTCGCCACTTCTTACTTCTCTCGGAGCA	3137
QY	2084	TGACGGGATGTGTGTACCGTCTCCCGTCTCTGCTACAAAGGGGAAAGCCATAGG	2143
DB	3138	TGTATTGGAACGGTAGTGTAAAGCCCTTTATCTGTTCAATTTTATGGGAATCGTTTCT	3197
QY	2144	ATCTCTCTTCCAAATAGGGGCTAGCCGATTGTTTCCAAACGATAGTCCCGCTTTTGGAG	2203

DB	3198	AAATCTTTGGCGGTATAGGATCTTAGTTCGTTAAAAATGGGATTTTTCCTTTAGAG	3257
QY	2204	TAGACGAGGATCATGCTCTTTTGGATCCGAAGCCCTTAGCGGAAAAGTTTTTGGGATTT	2263
DB	3258	TAAATAAATCATATCTTTTATATTGCCATAGCCCACTCTTTTAAAAATTTTAGGATTG	3317
QY	2264	GAACGATGCGGGGATGATGTTTAAAGATTTTCGCCGGGCAAGACCTCATCAAGATG	2323
DB	3318	CACTTTATGCGTGTGATTTTCAATTTATAAAATTTTGTATGCCAAATATTTTCACTAACATC	3377
QY	2324	AGCTTACCTCGAACCCGATTTCTCTCTATGTGAACGAAGATCAGTCTCTGAGTCCGCC	2383
DB	3378	ACCTTAACATAATGCCCTATCTTACATCTGTATGCACATAGATAGAGCTTTGTTCTGAA	3437
QY	2384	ATCAGCTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACCTCCCAAACTGAGGACCATCG	2443
DB	3438	AGCAATCTTTTAAGCAATACAGCGGTGTTTAAAAAATTTCTATAAAATCCATACCCACT	3497
QY	2444	AGGCTGTCATGTCAGCCCACTGACCGTTTTCGGCTGGCTGACGGTAGCAACGCGCATCT	2503
DB	3498	ACTTTGTCGTATAGCAATATCGCCATCTTGAATGCTAATGTGTTGCTGTCTACCA	3557
QY	2504	GTTTCATCGCGCCCAACGAGAACTGCTGGCCGCTTCCATAAGCGCGGTCAATATAGAAC	2563
DB	3558	TTT-----GTGATAGTAAAAATGTTGTTGTAGCAATGAGGGTCAATATATATC	3608
QY	2564	AACTGGACCTTCCC	2577
DB	3609	AAATCTATTTTTC	3622

RESULT 8

AE001444

LOCUS

DEFINITION Helicobacter pylori, strain J99 section 5 of 132 of the complete genome.

ACCESSION AE001444

VERSION AE001444.1

KEYWORDS GI:4154549

SOURCE Helicobacter pylori J99

ORGANISM Helicobacter pylori J99

REFERENCE 1 (bases 1 to 11131)

Alm,R.A., Ling,L.S., Moir,D.T., King,B.L., Brown,E.D., Doig,P.C.,

Smith,D.B., Noonan,B., Guild,B.C., deJonge,B.L., Carmel,G.,

Tummino,P.J., Caruso,A., Uria-Nickelsen,M., Mills,D.M., Ives,C.,

Gibson,R., Merberg,D., Mills,S.D., Jiang,Q., Taylor,D.S.,

Vovis,G.F. and Trust,T.J.

Genomic-sequence comparison of two unrelated isolates of the human

gastric pathogen Helicobacter pylori

Nature 397 (6715), 176-180 (1999)

99120557

9923682

2 (bases 1 to 11131)

King,B.L., Alm,R.A. and Trust,T.J.

Direct Submission

Submitted (12-JAN-1999) Astra Research Center Boston, 128 Sidney

Street, Cambridge, MA 02139, USA

Address all correspondence to: hp@arch.us.astro.com or Richard

A. Alm, Astra Research Center Boston, 128 Sidney Street, Cambridge,

MA, 02139. Lo-See L. Ling, Donald T. Moir, Douglas R. Smith,

Braydon C. Guild, Gilles Carmel, Anthony Caruso, Debra M. Mills,

Rene Gibson, and Gerald F. Vovis are with Genome Therapeutics

Corporation, 100 Beaver Street, Waltham, MA, 02453. Qin Jiang and

Diane E. Taylor are with the University of Alberta Department of

Medical Microbiology and Immunology, Edmonton, Alberta, Canada, T6G

2H7 and the Canadian Bacterial Diseases Network. All other authors

are with Astra Research Center Boston, 128 Sidney Street,

Cambridge, MA, 02139. Putative identifications, sequence

alignments, and name and sequence sequence capability are available

at ARCB's World Web site. (URL: <http://www.astro-boston.com/hpylori>).

FEATURES		Location/Qualifiers	
source		1. .11131	
		/organism="Helicobacter pylori J99"	
		/mol_type="genomic DNA"	
		/strain="J99"	
		/db_xref="taxon:85963"	
gene		complement (248. .1279)	
		/gene="jhp0045"	
CDS		complement (248. .1279)	
		/gene="jhp0045"	
		/codon_start=1	
		/transl_table=11	
		/product="putative TYPE II DNA MODIFICATION ENZYME (METHYLTRANSFERASE)"	
		/protein_id="AAD05627.1"	
		/db_xref="GI:4154550"	
		/translation="MQNLIQAEANAIALFLNDKLNKLGKIDLIYIDPPATNNHFTI	
		TNGRATITSNKNGDIAYSQKVGMDPWEFLKQRLVLLKELLSEQGSIVYHTDYKIGH	
		YVKVMDLEIFGIQNEIRIETPKCNPKFKRIGSGNIDKMLFYSGKKNFLENPKIP	
		YTPQLEKRRFPKIDKRRYITTFIHPAGEVESGEGCAFKGMLPKGRHWRDIAIL	
		ERWDKEGLIEYSNNNRKKIYALEQVKRQVDIWEFKDPQYPSYPTKNAQLLDLII	
		KTSNKKDSIVLDCGSGGTTLKSAFLIQRKFIGIDNSDLAIQACKNKLETITKDLFVS	
		QNFYDFIVF"	
		complement (1289. .2071)	
gene		/gene="jhp0045"	
		complement (1289. .2071)	
CDS		/gene="jhp0045"	
		/codon_start=1	
		/transl_table=11	
		/product="putative TYPE II restriction enzyme"	
		/protein_id="AAD05628.1"	
		/db_xref="GI:4154551"	
		/translation="MNVTKLSIEYANQRSVLDLFOVYPTIEGLREIDSKTMSNVE	
		YHFKKONLALITELLDLDPFIDKSYMAYLRKDSALERNPTIRNICRGLYEMGLN	
		KIFKCEPKEQNGIOPMKDQNNKSLGVPVVDNFANENDAILKASDNVMAEF	
		AKSHLNVHHHGLDFVAFKPKYIIGEAFLTFDFGHCNAQFNDAINTIETPNKAIK	
		VAILDGLVLYIESNNKMRKLDITVYRNHINSALVRLFLYQI"	
		complement (2499. .3989)	
gene		/gene="putp"	
		/note="synonym: jhp0047"	
		complement (2499. .3989)	
CDS		/gene="putp"	
		/note="similar to H. pylori 26695 gene HP0055"	
		/codon_start=1	
		/transl_table=11	
		/product="sodium/proline symporter"	
		/protein_id="AAD05631.1"	
		/db_xref="GI:4154554"	
		/translation="MGHYVLSPTIVTMFVVVSLMLYIGFYFQKNETTDFYLGDRS	
		MGPVISALSAGASDMGWLMLGLPCALYVGLINSIAIGLSGLGALINWVFAKRLRI	
		YTSVANSITISDYFETRFSDDKHILRLISAFVLIFFYFISGLVSGAKLFEATFG	
		IQYVVALSIGTLIIVSYTLGKYKAVCWTDLIQGLMMSALIVVPIVMIHLGGIGEG	
		IKIIRIPEKNSLQGSVVAIISLLAWGLGYGQPHILVREMSIIRDPVKATTI	
		GISMWVSLIGACVWGLLGVAIVHFKDLSLEDPEKIFIVSQLLFNWITGILLISAIL	
		AAVMSSTASSLIIVSSSTIAEDFYATIFNNAPQLKMTISKLSVLGVACIAFFISTDK	
		NASLISVSYAWAGFASGFSGVILFSLFWSRMTRIGAIAGLSGASTVILYDFKGSF	
		LDIYIVPGFVSAIVAVFSLFVSRSCTKEAFTMLKETIESLKH"	
		complement (4016. .7573)	
gene		/gene="putA"	
		/note="synonym: jhp0048"	
		complement (4016. .7573)	
CDS		/gene="putA"	
		/note="similar to H. pylori 26695 gene HP0056"	
		/codon_start=1	
		/transl_table=11	
		/product="proline/pyrroline-5-carboxylate dehydrogenase"	
		/protein_id="AAD05632.1"	
		/db_xref="GI:4154555"	
		/translation="MQKTIEDSLLELAKLQDSINHLSEQKAFHFMQKLLNNPENK	
		VMIIELMDFSRCDKAMQEFEMIEHVDLYKYSRIFSFSEFKWLMGLIFSGKMLPDMS	
		VPFFVNRSTDKAMVLDQESQLRILKRNKIKILNVNFIIEEVLGEEASARE	
		KYSALKKNYIOYISIKITTFISQINILDPYYSKEIVKRLDALVALALREEKQGMF	
		KFINLDMEEFDLELTVESPMESIAKFDLNAQIVLQAYIPDSYELKHLHAFSKERYL	
		KGLKDKIRFVKGANSEETIASMKDWALPTFFSSKQDTSNNKMLDFVLEGDNYKY	
		THIGAASHNIFEIAYVTRIHALNDPVLVLEHPSFEMLEGMSLQAQSELKEMHKLIIYA	
		PVCDEAHFNNAIAYVRLIDENSDNFMKAFNNLKVGTSWKDQEQRFLNSLKGITAT	
		LDNTHRTQDRNAKOTHTTYPNHSPKNSDITDFILKANREWAKVRDRMHNAPLIEL	
		VEIDGRFEDPNLTJLEVFDRTHHKKIASVHLADKEAILKALEVAKDSKRSQSFT	
		EIHALLISOTAOFLFRERRGLIGISALEVGTQPAETDAEVSEAIIDFLFVPSYSLVLOE	
		QNKTKTFPGGVVVIAPWNPFGVIGSVGTIAAPLAAGNRVIYKPSLSVSTGYKLCEC	
		FWDAGVPDALLIYFSGSDSISEHLKDESIRKFAITLTGEGDTAYKMLRANPTLALSAE	
		TGGKNATIVSKMADRQAIKNVIFAFNSGQKCSATSLVLEKEVIEDENPKTLID	
		ATLSISVGDFDFKNGIKALADKPNEKVIKAIDELKSYENVEIPASFVDDNPYLMKPS	
		IKYGTGKGFTHQTELFPTILSWKAKDLDEAIEIVNSTGYGLTSALSLEDRERWEYY	
		LERIBAGNIYINKPTTGAIQLROPFGVKSAVGFGRKVGIFNYITQFVNIHQEEDEN	
		NALKPLSEALRGLTKQGVDSHTHELKPAIMAKSYAVHYKHESQAKDYVKIRGEDN	
		LFSYKAVSVGRIYITEKDTLSLMDGLVALCLLSIQIPLTISIENERAKDLTFELCLK	
		ALQANAPIVYESLQRFSEKLNAFNRVIRIKDSLDLHQAQSLGMLVLAIAKPCLNGRF	
		ELLYHLRSVSISYHRYGNLGRVLRPTCHKSCAAEK"	
		7894. .8103	
		/gene="jhp0049"	
		7894. .8103	
		/gene="jhp0049"	
		/note="similar to H. pylori 26695 gene HP0057"	
		/codon_start=1	
		/transl_table=11	
		/product="putative"	
		/protein_id="AAD05633.1"	
		/db_xref="GI:4154556"	
		/translation="MKTIKNGIMIGTLGALLSLGCGSPDAQRFACILPKDHSKDASTK	
		KEAQYIPKGFDFPYSNLNHWDSF"	
		8431. .9093	
		/gene="jhp0050"	
		8431. .9093	
		/gene="jhp0050"	
		/note="similar to H. pylori 26695 gene HP0058"	
		/codon_start=1	
		/transl_table=11	
		/product="putative"	
		/protein_id="AAD05634.1"	
		/db_xref="GI:4154557"	
		/translation="MFKSSNDTQENPKDALKSTNFHEKHKTPTTETTELVEENKALTT	
		EXERLERENKLTADKENLTKEKTELQKOVNELKNSKVLENEKADWLREKENLTDDR	
		ENLTKEKTELTEKNKVLITTEKRLATEKENLTKEKTESOKVNELKNSKOVLENEKAD	
		LNNENTKLTDKTDTEKNQRLTTEKTELNNKITGLATEKERLAADKENLTRESQRK	
		PN"	
		9223. .9585	
		/gene="jhp0051"	
		9223. .9585	
		/gene="jhp0051"	
		/note="similar to H. pylori 26695 gene HP0058"	
		/codon_start=1	
		/transl_table=11	
		/product="putative"	
		/protein_id="AAD05635.1"	
		/db_xref="GI:4154558"	
		/translation="MEVKRLNFNFVATTHSAMDLIASVLSDSKVYLSFYNAQSOEL	
		GKRSKDGSKLAELFDLLPEYIKDSFELKEPSADYDCKLYKPEQNTSQMRRVVL	
		RGYKHNDKMYHTIVDMGS"	
		9586. .10569	
		/gene="jhp0052"	
		9586. .10569	
		/gene="jhp0052"	
		/note="similar to H. pylori 26695 gene HP0059"	
		/codon_start=1	
		/transl_table=11	
		/product="putative"	
		/protein_id="AAD05629.1"	
		/db_xref="GI:4154552"	
		/translation="MGTLEKWFQSOIREELREARISELENTTELLREREYLAETS	
		ELKDQALQKNDKLPITKDKLTKEKTELPAENESLSVYKISGLEHNDQLMNNKL	
		TEKAEKTEKTLAKENTRLAARDRLTEERELTEKERLAKRENTTELTHKITELTK	
		ENKALPTEDKLNHQTVAUTNERDSUEQERAKQDAGHLEKRCNTLNKENORLTDKL	
		KQLSEAKSLENTNQLQALNSNVQAQAKETIAIEKSELEFARLKSLEGMEEK	
		SOLDLHNRLASANDLQKRNKLEBENTALKERVVDGLNEQLSKLPQKPFQ"	
		10580. .11011	

```
CDs
/genes="jhp0053"
10580..11011
/genes="jhp0053"
/notes="similar to H. pylori 26695 gene HP0060"
/codon_start=1
/transl_table=11
/product="putative"
/protein_id="AAD05630.1"
/db_xref="GI:4154553"
/translation="MYKESLPLCHPLVEXKLCBLSLNNAFNEKSEPNPKDFVSE
EDIKFKQALRVFHALVNNETSIRYJSDENQFLDLAQAKKITNEIKLEILVSDS
DVEDFEKLKMLNVDVAVGLSKYSQNLDLDDGGHDL"

ORIGIN
Query Match 3.2%; Score 185.6; DB 1; Length 11131;
Best Local Similarity 52.0%; Pred. No. 5.4e-29;
Matches 475; Conservative 0; Mismatches 424; Indels 15; Gaps 2;

QY 1664 ACCCTTCTAGTAGAGCTTTGACCGCTCCAGGAGGCATCTATGCGGATCGCGGC 1723
DB 309 AGCTTGTTTTGCAAGCTTGATAGCCAAATCGGAATTATCAATGCTTATAAATTTGCT 368
QY 1724 TTAAAGGGGTGAGCTATAACGCTAGTACCGGACCTCGAAGGATCGAGCACTAAA 1783
DB 369 TGCATAAAAACGCAGATTTAAGGTGTTCAGAACCAACAAACATCTAAACAATA 428
QY 1784 TCCCTCTGTTACTCTCTGTTGGACGATGAGCTTGAGCATGTCCAGATTTTTCGGTG 1843
DB 429 CTATCTTTATAGAGAGGTTTTAATGATTAACTCTAATTAAGTCTAATTAAGAGCTTTTCTGTA 488
QY 1844 GGTATCGGGGTACGAGGATCTTGAACCTGCTGCAACGCTCTGAGGCTTCTCCCTTC 1903
DB 489 GGTAGCTTGGATATTGTGGCTCTTTAAATTCAGATGCTTGGACTTTTGGC---- 544
QY 1904 TTCAGGCGATCCGAGCGTAAACTTCTTCCCGGCGCACCCGCTCTTTGACGAGCAATA 1963
DB 545 --NACTTGTTCTAAGGCATAAATTTTTTCTAGGGTTATTATTGTAGAACTCAATC 602
QY 1964 AGCCCTTAGCGTCTAGCTGCTCAAGCTTCTCGGGGGATAGGCGCAATGCGGTCGAGCA 2023
DB 603 AAACCTTCTTTATCCCAAGCGCTCAAGTGTGGAATATCATAGTGGCAATGCGCGCTTTT 662
QY 2024 GGGGGAAGTATTCCTCGCAAGSCCTTCGGTAGGSCCATCTCTGTTTCTCCAGAGCA 2083
DB 663 GGAGGTAGCAATCTTAATGCTTTAGAACATTCGCCACTTTCACCTTCCTCTCGAGCA 722
QY 2084 TGACGGGATTTGTTGTTGATACCGTTCCCGTTCTCGTCTACAAAGGGGAAAGCCCTAGCG 2143
DB 723 TGTATTGGAACGGTAGTGTAAAGCGCTTTTATCTTGTCAATTTTAGGGAATCGTTTTCT 782
QY 2144 ATCTCTCTTCGAATAGGCGGTAGCGGATTCGTTCCAAACGTAGTCCCGGCTTTTGAG 2203
DB 783 AAATCTGTGGCGTATAGGGATCTTAGGTTCTGTTAAATAATGGGATTTTTTCTTTAGAG 842
QY 2204 TAGACGAGATCATGTCTCTTTTCCGATCCGAAGCGCTTACGGGAAAGTTTTTGGGATT 2263
DB 843 TAAATAAATATATATCTTTATGTTTACCATAGCTATCTCTTTAAATTTTAGGATTG 902
QY 2264 GAAGCGATGGGCGATATGTTTAACGAAGTTTCGCCGCGCAAGACCTCATCAGGATG 2323
DB 903 CACTTATGCGTGTGATTTCACTTCTAAATAATTTGTATGCCAAATATTTTCATCTAACAT 962
QY 2324 AGTTTCACTCGAACCCTGATTTCTCGTCTATGTGAAACGAAGATCAGTCCCTGAGTCCGCC 2383
DB 963 ACCTTGACATAATGCCCTATCTGTAATCTGTATGCACATAGATAGAGCTTTGTTGAA 1022
QY 2384 ATCAGTCTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAATCCACAACTGAGGACATCG 2443
DB 1023 AGCAATCTTTTAGCAATACCGAGCGCTGTTTAAATAATTTCCATAAATCCATACCCACT 1082
QY 2444 AGGGTCTCATCGTAGCCCACTGACCGTTTTTGGGCTGCTCAGCGTAGCAACGGATCT 2503
DB 1083 ACTTTATCACTATAGCAATATCGCAATCTTTAGAAATGCTTAATTTGTGTTGCTCTACA 1142
```

```
QY 2504 GTTTCATCGCGCCACAGAGAACTGCTGCCGTTCCATAAGCGCGGTCAATATAGACC 2563
DB 1143 TTT-----GTGATAGTAAATGATGTTTGTAGCAATGAGGGTCAATATATATC 1193
QY 2564 AACTGGACCTTCCC 2577
DB 1194 AAATCTATTTTCC 1207

RESULT 9
AF328909 4931 bp DNA linear BCT 12-NOV-2002
LOCUS Helicobacter pylori strain CR2 putative cytosine-specific DNA
DEFINITION methyltransferase, putative HP0052-like protein, type II DNA
modification enzyme, and type II restriction enzyme genes, complete
cde
ACCESSION AF328909 GI:17225495
VERSION AF328909.1
KEYWORDS Helicobacter pylori
SOURCE Helicobacter pylori
ORGANISM Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
Helicobacteraceae; Helicobacter.
REFERENCE Chanto,G., Occhialini,A., Gras,N., Alm,R.A., Megraud,F. and
AUTHORS Marais,A.
TITLE Identification of strain-specific genes located outside the
JOURNAL plasticity zone in nine clinical isolates of Helicobacter pylori
PUBMED Microbiology 148 (Pt 11), 3671-3680 (2002)
AUTHORS 12427957
TITLE 2 (bases 1 to 4931)
REFERENCE Chanto,G., Occhialini,A., Gras,N., Megraud,F. and Marais,A.
AUTHORS Direct Submission
JOURNAL Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES
Location/Qualifiers
1..4931
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="CR2"
/specific_host="Homo sapiens"
/db_xref="taxon:210"
/map="region P between JHP43/HP50 and JHP47/HP55"
/notes="isolated from patient with gastric carcinoma"
complement(1..1086)
/notes="putative HP0051-like protein"
/codon_start=1
/transl_table=11
/product="putative cytosine-specific DNA
methyltransferase"
/protein_id="AAL37432.1"
/db_xref="GI:17225496"
/translation="MNYKILDFPCGAGGSAGLYVKEDALIGLDCDQALITFENN
HKNAIGVGDITQTEIKKVIELAKLEINMIGPPCCQGFNSKNGKLGKDPNPLF
LEYIEVKAIPKFIENYKLIISCAKGYFLEEIKERLNALGYQSYQLNAKDYGS
PQNRKAFIVGASRFDNLLPSPVNVQDAISDLAYLCNEGAFSDYINPIQSS
YQALMRKDSKLYNHOATNHSQAELKLNKEQKCLPKNLHGKQOFKSTWGRLN
WKLISITDTPRPTPSGNTSHPELHRSITPREAARIQSFSDDYIFYGNKTSVCKQIG
NAVPPILALGAKLAKSARNUTLSR"
complement(1083..2309)
/codon_start=1
/transl_table=11
/product="putative HP0052-like protein"
/protein_id="AAL37433.1"
/db_xref="GI:17225497"
/translation="MLVSRPLNDDIDPFLGVLLSRFOJNGCIYVCVSKASKFVPD
YKSKAQVNLALNTLSVHPVWQSNQSVTKIGTFVLEKDLHDDENAFYKGLMLLI
INDFFNRGHSMTPNQKFLSGFFESRGSIDTQNFELTSDYFFHSPLEFNKGLYLI
FNPSEALNPNFRELQPNQNAQRIYLNWYLLHIGLGNFYKARIAHVFKT
FTLYDGTIYKLSYPTTKYHGNSFTERAHFYLNWVQDLDKSEIKLEQQLGWIOKS
EERKRSKLIINFYRISITLVNVCACCGDYDKERSFISLPLYKTQNPNSYTYEIHVV
SLGKNELDVLENLAKLCTCTCHALKKGSEEFQKRLRNILNHNKDNLEPAQLRFE
```

```

TDDFLTLIDKIHESLK"
complement (2677..3708)
/function="methyltransferase"
/note="JHP45-like protein"
/codon_start=1
/transl_table=11
/product="type II DNA modification enzyme"
/protein_id="AAL37434.1"
/db_xref="GI:17225498"
/translation="WQNLIOAENAIALLFLNDKNLKGIDLIYIDPPFATNNHFTI
TNGRATTISNKGIDIAYSKVGMDPIELFKQLVLLKEILLSEQSGSIYVHTDCKIGH
YKVLMDIEI RGIQNFREI TRI KGNPKNFKMVGNI KDMILFVSKGNI FNEPKIP
YTPDLERPKIDKRRYTVPIHAPGEVSEGCSEKAFKGMPLPKGRHRTDIATL
ERWKEGLIEVSNNDNRKRIYALEOVGKRVQDIWEKDPQYSPYFTEKNAQLDLII
KTSNKDSIVLDCFCGSGTTLKSAFLIQRKFIDGIDNSNLAIQACKNKLETTITKDLFVS
QNFYDFLVF"
complement (3718..4503)
/note="JHP46-like protein"
/codon_start=1
/transl_table=11
/product="type II restriction enzyme"
/protein_id="AAL37435.1"
/db_xref="GI:17225499"
/translation="WVNYWTKLSIEYANQSRSLDLFQVYPTIPEGUREIDSKIWSNI
BYFKQKNLALITELNLIDLPFKIDKSYMLKEDKQALERNPTNIRICRLYEMGL
NKIPEKCEPKEIQRIGMPKFNKNSLGVPEVDLNDPIANENDAI LRASDNIMAE
PAKSHLNYHHKGLDFVARFNKYIIGAEKELTDFGGHQAQENDAISTIEAPNIKAI
KVALIDGVLYIESNNKRLDITTYKININISALVJREFLYQI"

ORIGIN
Query Match          3.1%; Score 184; DB 1; Length 4931;
Best Local Similarity 51.9%; Pred. No. 1.2e-28;
Matches 474; Conservative 0; Mismatches 425; Indels 15; Gaps 2;

QY 1664 ACCCTTCTAGTGAAGCTTTGACCCCTCCAGAGGCACTATCGCATGGATCGCGC 1723
DB 2738 AGCTTGTTTTGCAAGCTTGGATGCTAAATTTGGAATTTATCAATGCCTATAAATTTTCGT 2797
QY 1724 TTATAGAGGGGTAGGCTATAAGCTAGTACGGAGCTCGGAAGGGATCGAGCACTAAA 1783
DB 2798 TGCAATAAAACCGAGATTTTAAGTTGTTCCAGNACCACAAAACATCTAAACAATA 2857
QY 1784 TCCCTCTCTACTCCCTGTTTGGACGATGAGCTTGACATGTCAGATTTTTCGGTG 1843
DB 2858 CTATCTTTATAGAAGAGTTTAAATGATTAAGTCTAATAATGAGCGTTTTCCTGTA 2917
QY 1844 GGGTATCGGGTACGGAGGATCCTTGAACTGCCAAGCTCTGGAGCTTCTCCCTTC 1903
DB 2918 GGGTAGCTTGGATATGTGGGGTCTTAAATTCCTCCAGATGCTTGGACTCTTTGCC---- 2973
QY 1904 TTCAGGCGATCCGAGCGTAAACTTCTTCGCGCGCACCCCGTCTTTTGACCAGACAATA 1963
DB 2974 --AAGTTGTTCTAAGGCATAAATTTTTTTCTAGGGTTATTATTGTTAGATACTCAATC 3031
QY 1964 AGCCCTTAGCGGTCTAGCTCGTCAAGCTTCTCGGGGGATAGCGCAATGCGCTCCAGGA 2023
DB 3032 AAACCTTCTTTATCCCAACGCTCAAGTGTGGCAATPATCAGTGCAGCAATCGCGCCCTTT 3091
QY 2024 GGGGGAAGTATTCTTCGCGCAAGGCTCTCGGTAGGGCCATCCTGGTTCTTCCAGGAGCA 2083
DB 3092 GGAGTAGCATACCTTTAAATGCTTTAGNACATTCGCCACATTTCCACTTCTCTCTGGAGCA 3151
QY 2084 TCGAGGGAATGGTGGTGACCGTTCCCGGTTCTCGTCTACAAAGGGGAAAGCCTTAGCG 2143
DB 3152 TGTATTGAAACGGTAGTGTACGGCTTTGTCTTGTCAATTTTAGGGAATCGTTTCT 3211
QY 2144 ATCTCTCTTCCGATAGGGCTACCGATTCGTTCCAAAGCTAGTCCCGGTTTGGAG 2203
DB 3212 AAATCTTGTGGGTATAAGGATCTTAGTTCGTTAAATAAGGATTTTTTCCCTTTAGAG 3271
QY 2204 TAGACGAGGATCATGCTCTTTTTCGATCCGAGGCTTACCGGAAAAGCTTTTGCGATTT 2263
DB 3272 TAAATAAAATCATATCTTTTATGTTGCCATAGCCCATCTTTTAAATAATTTTAGGATTG 3331

CDS
2264 GAAGCATCGGGGAGATATGTTAAGAGATTTCCCGCCGCAAGACCTCATCAAGATG 2323
3332 CACTTTATCGGTGTGATTTTCATTTCTAAATAATTTGTATGCCAAATATTTTCATCTAACATC 3391
2324 AGCTTCACCTCGAACCCGTAATTTCTGCTATGTGAACGAAGATCAGTCTCCTGAGTCCGCC 2383
3392 ACCTTAACATAATGCCCTATCTTACAACTGTATGCACATAGATAGAGCTTGTCTTGAA 3451
2384 ATCAGTCTCCTGAGAGTATCAAGCGCTCCCTCAGAACTCCACAAATGAGAGCATCG 2443
3452 AGCAATTTCTTTAAGCAATACCAAGCGCTGTTTAAATAATTTTATAAAATCCATACCCATC 3511
2444 AGGCTCTCATCTAGCCCACTGACCGTTTTTGGCTGGCTGACGCTAGCAACGCGATCT 2503
3512 ACTTTGTCACTATAGCAATATCGCCATCTTAGAATTCCTAATTTGTGTTGCTCTGCCA 3571
2504 GTTTCATCGCCGCCAACGAGAACTGCTGCGCGTTTCCATAAGCGGGTCAATATAGACC 2563
3572 TTT-----GTGATAGTAAATGATTGTTTGTAGCAAAATCGAGGGTCAATATATATC 3622
2564 AACTGACCTTCCC 2577
3623 AAATCTATTTTCCC 3636

RESULT 10
AF328912 3172 bp DNA linear BCT 12-NOV-2002
LOCUS Helicobacter pylori strain CR14 JHP44-like protein, type II DNA
DEFINITION modification enzyme, and type II restriction enzyme genes, complete
cds.
ACCESSION AF328912
VERSION AF328912.1 GI:17225508
KEYWORDS Helicobacter pylori
SOURCE Helicobacter pylori
ORGANISM Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.
REFERENCE 1 (bases 1 to 3172)
AUTHORS Chanto,G., Occhialini,A., Gras,N., Alm,R.A., Megraud,F. and
Marais,A.
TITLE Identification of strain-specific genes located outside the
plasticity zone in nine clinical isolates of Helicobacter pylori
JOURNAL Microbiology 148 (Pt 11), 3671-3680 (2002)
PUBMED 12427957
REFERENCE 2 (bases 1 to 3172)
AUTHORS Chanto,G., Occhialini,A., Gras,N., Megraud,F. and Marais,A.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES
source
1..3172
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="CR14"
/specific_host="Homo sapiens"
/db_xref="taxon:210"
/map="region F between JHP43/HP50 and JHP47/HP55"
/note="isolated from patient with duodenal ulcer"
complement(1..558)
/function="methyltransferase"
/note="type II DNA modification enzyme"
/codon_start=1
/transl_table=11
/product="JHP44-like protein"
/protein_id="AAL37442.1"
/db_xref="GI:17225509"
/translation="MLVSEFLNAIDPNLGVLLSRPQIKNGCTIGVCSYRASKFTSGY
ESKARVLMALNTLSHQIWNQCSYTKGTFFVLENDLHLDENASYKLLNLLI
DNDFNRHLVTPSNGTNSHPHLRSITTPREATRIQSFDDTIYFNKTSVCKQIGNA
VPELLALGAILKARSNDTNLSR"
complement(921..1868)
CDS

```

```
/function="methyltransferase"
/notes="JHP45-like protein"
/transl_table=1
/codon_start=1
/product="type II DNA modification enzyme"
/protein_id="AAL37443.1"
/db_xref="GI:17225510"
/translation="MIVIDPPANTNNHFTTNGRATTSNKNNGDIAYSDKVGWDFI
EFLKQLRLVLELLSQSGIYVHTDYKIGHYKVMLEDFIQNFRNITRIKKNPKN
FRIGVGNKIMLFFSKGNPLFNEPKIPYQDLERFPKIDKDKKRIYTPVPHAP
GEVSEGCDAKMGMLPPKGRHRTDIATLERWDKLEGLYEYNNNNPKRIYAEQVG
KRVQDIWEFKDPQPSYPTKNAQLDLIIKTSNNKDSIVLDFCGSGTTLKSAFLQ
RPFIDNGSLGAIKACKNLETKIDLFISQNFYDLVF"
complement(1963..2652)
/notes="JHP46-like protein"
/codon_start=1
/transl_table=1
/product="type II restriction enzyme"
/protein_id="AAL37444.1"
/db_xref="GI:17225511"
/translation="MREIDSKTWSNIEYHFKODNLITLITLNLDLPIKDSYIAYL
KDKSLERNPTINICGLYEMGVNKKIFKCSBKETNROIGPMFKDMLNKKSLGV
EPVDLNDFTANEDAILKASDNVMAEPFAKSHLNYHHKGLDFVARNFKYIIGEKFL
TFPGHQNAQFNDAISTIEAPNIKAIKVAILDGVLYIESNNKORKLLDITTYANVHIMS
ALVIREFLYQI"

ORIGIN
Query Match 3.1%; Score 182.4; DB 1; Length 3172;
Best Local Similarity 51.8%; Pred. No. 2.8e-28;
Matches 473; Conservative 0; Mismatches 426; Indels 15; Gaps 2;

QY 1664 ACCCTTCTAGTGAAGCTTTGACCGCTCCCGAGGAGCATCTATCGCATGGATCGCGC 1723
DB 982 AGCTTGTCTTTGCAAGCTTTGATAGCTAAACCGAATTATCAATGCTATAAATTTGCT 1041
QY 1724 TTAAAGGGGTGAGGCTATAGCGTAGTACCGGACCTCGGAGGGATCGAGCACTAA 1783
DB 1042 TGCAATAAAACCGAGATTTAAGGTGTTCAGAACACCAAAACAACTATAAACATG 1101
QY 1784 TCCCTCTGTTACTCCCTGTTGGAGCATGAGCTTGAGCATGCTCCAGATTTTCTCGGTG 1843
DB 1102 CTATCTTTATAGAGAGGTTTAAATGATTAAATCTAAATTTAGCGGTTTCTCTGTA 1161
QY 1844 GGTATCGCGGGTACGAGATCTTGAACGTGCAACAGCTCTGAGCTTCTCCCTTC 1903
DB 1162 GGGTAGCTTGGATATTGCGGTCTTTAAATCCAGATGCTTGGACTTTTGGC---- 1217
QY 1904 TTCAAGCGATCCGAGCGTAAACTTTCTTCCGCGGCAACCCCTTCTTTCAGCAGACAATA 1963
DB 1218 --AACTTGTCTAAGCATATAATTTTCTTAGGTTATTATTGTAGATCTCAATC 1275
QY 1964 AGCCCTGAGCGTCTAGCTGTGCAAGCTTCTCCGGGGATAGCGCAATGCGCTCAAGA 2023
DB 1276 AAACCTTCTTTATCCCAACGCTCAAGTGTGGAATATCAAGTGGCAATGCGCGCTTTT 1335
QY 2024 GGGGGAAGTATCTCTCCCAAGCGCTTCCGTTAGGCGCATCTCTGTTTCTCCAGGACA 2083
DB 1336 GGAGGTAGCATACCTTTAAATGCTTTAGAACATTCGCCACTTCTACTTCTCCTCGAGCA 1395
QY 2084 TGCAGGGGATGTGGTGTACCGTCTCCCGTTCTCGTCTPACAAAGGGGAAAGCCATCGC 2143
DB 1396 TGATTGGAACGGTAGTGTAAAGCGCTTTTGTCTTGTCAATTTAGGAATCGTTTTTCT 1455
QY 2144 ATCTCTCTCTCGAATAGGGGTAGCGGATTCGTTCCAAACGTAGTCCCGCGTTTGGAG 2203
DB 1456 AAATCTGTGGCGGTATAAGGGATCTTAGGTTCTGTTAAATAATGGGATTTTTCCTTAGAG 1515
QY 2204 TAGACAGGATCATGTCTCTTTTCGATCCGAGGCGCTTACGGGAAAGTTTTTGGGATTT 2263
DB 1516 TAAATAAAATCATATCTTTATATPACCATAGCCTATCTCTTTTAAATTTTAGGATTG 1575
QY 2264 GAAGCGATCGGGCGATATGTTTAAAGAGTTTCGCCGCCGAAACACCTCATCAGGATG 2323
```

```
DB 1576 CACTTTATCGGTGTGATTTTCATTTCTTAAATTTTGTATGCAAAATATTTTCATCTAACATT 1635
QY 2324 AGCTTACCTCGAACCCTGATTTCTGCTATATGTGACGAAGATCAGTCTGTGATCGGCC 2383
DB 1636 ACCTTAACATAATGTCCTATCTTATAATCTGTATGCACATAGATAGAGCCTTGTCTTGAA 1695
QY 2384 ATCAGCTCCCTGAGAACTATCAAGCGCTCCCTCAGAACTCCCAAACTGAGGACCATCG 2443
DB 1596 AGCAATTTCTTAAGCATACTAGCGTGTCTTTTAAATTTCTATAAATCATACCCACT 1755
QY 2444 AGGCTGTATCGTAGCCCAACTGACCGCTTTTGGCTGGCTGACGGTAGCAACCGCATCT 2503
DB 1756 ACTTTGCTGCTATAAGCAATATGCCATCTTGAATTTGTAATTTGTTGGTGTCTTCGCCA 1815
QY 2504 GTTTTCATCGCGCCACGAGAACTGCTGCCCGTTTCCATAAGCGGTCATATATAGACC 2563
DB 1816 TTT-----GTGATAGTAAATGATTTGTTAGCAATGAGGGTCAATATATATC 1866
QY 2564 AACTGGACCTTCCC 2577
DB 1867 AAATCTATTTTCC 1880

RESULT 11
AF328911
LOCUS
DEFINITION
Helicobacter pylori strain CR9 type II DNA modification enzyme,
type II DNA modification enzyme, and type II restriction enzyme
genes, complete cds.
ACCESSION
AF328911
VERSION
AF328911.1 GI:17225504
KEYWORDS
Helicobacter pylori
SOURCE
Helicobacter pylori
ORGANISM
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.
REFERENCE
1 (bases 1 to 3261)
Chanto, G., Occhialini, A., Gras, N., Alm, R.A., Megraud, F. and
Marais, A.
Identification of strain-specific genes located outside the
plasticity zone in nine clinical isolates of Helicobacter pylori
Microbiology 148 (Pt 11), 3671-3680 (2002)
12427957
REFERENCE
2 (bases 1 to 3261)
Chanto, G., Occhialini, A., Gras, N., Megraud, F. and Marais, A.
Direct Submission
Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
Location/Qualifiers
FEATURES
source
1..3261
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="CR9"
/specific_host="Homo sapiens"
/db_xref="taxon:210"
/map="region P between JHP43/HP50 and JHP47/HP55"
/notes="isolated from patient with gastric carcinoma"
complement(1..558)
/functions="methyltransferase"
/notes="JHP44-like protein"
/codon_start=1
/transl_table=1
/product="type II DNA modification enzyme"
/protein_id="AAL37439.1"
/db_xref="GI:17225505"
/translations="MDVSRFLNADPFLNGLVLSRFOIKNGCIYGVCSYKASKFTSSY
EESKAVLNALNLSKHQWQSNQSVTKVKGTFVILENDLNLDNATFKLLNLLI
DNDFRPHLVTPTSGNTSHPELHRSITPREARIQSFSNDYIFYGNKTSVCKQIGNA
VPFLALGKALIKSTRNDTLNR"
complement(1008..2039)
/functions="methyltransferase"
/notes="JHP45-like protein"
/codon_start=1
```

```

/transl_table=11
/product="type II DNA modification enzyme"
/protein_id="AAL37440.1"
/db_xref="GI:17225506"
/translation="MQNLLIQAESIALFLNDRKLNKLGKIDLIYIDPPATNNHFTI
NGKATTSKNGDIAYSQVGMDFIEFKORLVLLKELLSEQSIYVHTDKIGH
YVKWMLDEIPIQIONFRNEITRIKPNQFKRIYGNIKOMILFYSKGNPIFNEKIP
YTPDLKRPKIDKORRYTTPVIAHAPGEVSEGSKAFKGMPLPKGRHRTDIALI
ERWDKEGLIYSNNNRPKKIYALEOAGKAVQIDIMFKPKQPSYPTKNAQLDLII
KTSNKRDSIVLDCFCGSGGTTIKSAFLQRFIIGIDNSDLAIQAACKNLEITIKDLFVS
QNFYDFLIVF"
complement (2049..2834)
/notes="JHP46-like protein"
/codon_start=1
/transl_table=11
/product="type II restriction enzyme"
/protein_id="AAL37441.1"
/db_xref="GI:17225507"
/translation="MVNYTKLSIEYANQSRISLDDLPQVYTIPEGLREIDSKIWNI
EYFKQDNLALITELLNLDLFTKDSYIAYLKDKSLERNPTINRICRLYEMGL
NKIPEKSEPTNRQIGPMFKWLNKSLGVPDLNDFIANENDAILKASDNVMAE
PAKSHLVHHKGLDFVARNKYIIGEAFLTDFGCHQNAQFNDALSTIKADNIKAI
KVALDGVLYIESNNKRLKLLDTYRNYHMSALVLRFLYQI"
Query Match 3.1%; Score 180.8; DB 1; Length 3261;
Best Local Similarity 51.6%; Pred. No. 6.1e-28;
Matches 472; Conservative 0; Mismatches 427; Indels 15; Gaps 2;

ORIGIN
QY 1664 ACCCTCTAGTGAAGGCTTACCGCTCCAGGAGCACTATGCCGATCGCGGC 1723
DB 1069 AGCTTGTTTTGCAAGCTTGATGCCAAATCGGAATTATCAATGCCCTATAATTTTCGT 1128
QY 1724 TTTAAGAGGGGTGAGGCTATAGGCTAGTACCGGAGCTGCGAAGGGATCGAGCACTAAA 1783
DB 1129 TGCATAAAACGCGAGATTTTAAAGTTGTTCCAGAACCAACAAATCTAAACAATA 1188
QY 1784 TCCCTCCGTACTCCCTGTTTGACGATGAGCTTGACATGTCAGATTTTTCGGTG 1843
DB 1189 CTATCTTTATAGAGAGGTTTAAATGATTAACTAATAATGAGCGTTTTCCTGTG 1248
QY 1844 GGGTATCGGGTACGAGGATCCCTTGAACGCCAACGCTCGGAGCTTCTCCCTTC 1903
DB 1249 GGGTAGCTTGATATGTTGGTCTTTAAATCCAGATGCTTGGACTCTTTCCAGCT 1308
QY 1904 TTCAGCGATCCGAGCGTAACTTTCTCCGCGCACCCCGTTCTTTGACGACAAATA 1963
DB 1309 T-----GTTCTAAGGCATAAATTTTTTTCTAGGGTATTATTCTTAGAATACTCAATC 1362
QY 1964 AGCCCTTGAGGCTAGCTCGTCAAGCTTCTCGGGGATAGCGCAATGCGTCCAGGA 2023
DB 1363 AAACCTCTTTATCCCAAGCTCAAGTGTGGCAATATCAGTGGCGCAATGCGGCTTTT 1422
QY 2024 GGGGAAGTATTCCTCGCAAGGCTCTCCGTTAGGGCCATCCTTGTTTCTCCAGGAGCA 2083
DB 1423 GGAGTAGCATACCTTTAAATGCTTTAGAGCATTCGCCACTTCTACTTCTCTCGGAGCA 1482
QY 2084 TGCAGGATATGTTGGTGTACGGTTCCCGGTTCTCGTCTCAAGAGGGAAGGCTAGCG 2143
DB 1483 TGTATTGGAACGGTAGTGAAGCGCTTTATTCCTTGTCAATTTTAGGGAATCGTTTTC 1542
QY 2144 ATCTCTCTTCCGAATAGGGGTAGCGATCGTTCACAAAGTACTGCGCGGTTTGGAG 2203
DB 1543 AAATCTTGGGTATAGGAATCTTAGCTCGTTAAANAUGGATTTTTTCCITTAGAA 1602
QY 2204 TAGACGAGATCATGCTTTTGGCATCGGAAGGCTTACGGGAAAAGTTTTCGGATTT 2263
DB 1603 TAAATAAAATCATATCTTTATGTTAGCATAGCCTATCCTTTTAAAAATTTTAGGATTG 1662
QY 2264 GAAGCGATCGGGCATATGTTTACGAGTTTCCCGGCCAACACCTCATCAAGGATG 2323
DB 1663 CACTTTATACGTTGATTTCACTTTCTTAAAAATTTTGTATGCCAAATATTTCATCTCAACT 1722

```

```

QY 2324 AGCTTCACCTCGAACCCGCTATTTCCTGCTATGTGAACGAAGATCAGTCTGAGTCCGCC 2383
DB 1723 ACCTTAACATAATGCTCTTATAAATCTGTATGCACATAGATAGAGCCTTGTCTGAA 1782
QY 2384 ATCAGCTCCTCAGAAAGTATCAAGCGCTCCCTCAGGAATCCACAACTGAGGACCATCG 2443
DB 1783 AGCAATCTTTTAAGCAATACATAGGCGTTCCTTTAAAAATTTCTATAAATCCATACCCACT 1842
QY 2444 AGGGTGTCACTGATAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGCGATCT 2503
DB 1843 ACTTTGTGCTATAGCAATATCGCAATTCITAGAAATGCTAAATTTGTTGTTCTCTACCA 1902
QY 2504 GTTTCATCCCGCCCAACGAGAACTGCTGCGCGGTTCCATAAGCGGCTCAATATAGACC 2563
DB 1903 TTT-----GTGATAGTAAATGATTTGTTTAGCAATGAGGCGTCAATATATATC 1953
QY 2564 AACTGACCTTCCC 2577
DB 1954 AATCTATTTTCCC 1967

RESULT 12
AF328915
LOCUS
DEFINITION
Helicobacter pylori strain CR29 cytosine-specific DNA
methylation transferase, putative HP0052-like protein, type II DNA
modification enzyme, and type II restriction enzyme genes, complete
cnds.
ACCESSION
AF328915.1 GI:17225516
VERSION
AF328915
KEYWORDS
Helicobacter pylori
SOURCE
Helicobacter pylori
ORGANISM
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.
REFERENCE
1 (bases 1 to 5001)
Chanto,G., Occhialini,A., Gras,N., Alm,R.A., Megraud,F. and
Marais,A.
TITLE
Identification of strain-specific genes located outside the
plasticity zone in nine clinical isolates of Helicobacter pylori
JOURNAL
Microbiology 148 (Pt 11), 3671-3680 (2002)
PUBMED
12427957
REFERENCE
2 (bases 1 to 5001)
Chanto,G., Occhialini,A., Gras,N., Megraud,F. and Marais,A.
Direct Submission
TITLE
Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
JOURNAL
FEATURES
Location/Qualifiers
1..5001
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="CR29"
/species_host="Homo sapiens"
/db_xref="taxon:210"
/notes="region F between JHP43/HP50 and JHP47/HP55"
/notes="isolated from patient with chronic gastritis"
complement(1..1086)
/notes="HP0051-like protein"
/codon_start=1
/transl_table=11
/product="cytosine-specific DNA methyltransferase"
/protein_id="AAL37447.1"
/db_xref="GI:17225517"
/translation="VNYKILDFCGAGGFSAGLECLKEFDALIGLDCDQALITFENN
HKNATGICDITQIEIKKVIKLAOLNMIIGPPCGFNKGNLGLKDPNLF
LEVILVKAIKPEIFIENKNIISCAGYFLEIEIKERNALGYQLYLNAKDYGV
PQNRKAFVIGASRFSFDNLLPSSQVNIQDAISDLVLCNSGAFESDYLNPIQSS
YQALMRKSPKLYNHQAALSKLKNKEQKCLPKNLHKGQCFKSTWGLN
NWKISPTIDTFDTPSNGNSHPELHRSITPREATRIQSDDIIFYGNKTSVCKQIG
NAVPLIALALGKAILKSARNDNLSR"
complement(1083..2312)
/codon_start=1
/transl_table=11

```


/product="putative HP0052-like protein"

/protein_id="AAL37448.1"

/db_xref="GI:17225518"

/translation="MLVSRFLNAIDPNFLGVLLSRFOIKNGICVGVCSKLSKPTSPY
EESKARVLAVALNLSKHQIQSQNESVTKIKGTFVLENDLHLENDAFYKLLNLI
DNDFFRSHLNSFNQRFUSQFFESRGSIDTQRNFLTSDHFFSPDFNFKPLI
DHTFIFSEALNFWRELQPEYQVQIKRQAQFRIYLDWYHISLFPNFKAKIAEYFK
TLVANGYKIIISYPRTKYHNSFIERAHFLKXNYQDLDDKSIKLEQLGFIQD
SEFRDYSKIIISYRSLSTNVNACDYDDIKERSFISLPLQITQSRSDSYTEIHV
ISLQKELDLLENLAKLPTCHRALRKASAEFQKRLIRKLRKNDLNULEFAKLR
ETDDFLTDKXVESLK"

complement (2752..3783)

/function="methyltransferase"

/notes="JHP45-like protein"

/codon_start=1

/transl_table=11

/product="type II DNA modification enzyme"

/protein_id="AAL37449.1"

/db_xref="GI:17225519"

/translation="MQLNLIQAEALFLFLNDKLNKIDLIIVDPPFATNNHFTI
TNGRATISNKNQDIAYSDKVGMDFIEFLKQRLVLLKELLSEQGSIVHTDKIGH
YVKWLEDFIQNRENEITRIKCNPNFKMGYGNIKMILFYSGKNPIFNPEKIP
YTPQDLKEFPKIDKRRITVPIHAPGEVSEGSKAFKMLPKGRHWRITDIATL
ERWDKGLIYENNNPRKIVALEQVGRVODIWEFKDPOVPSYPTKNAQLDLII
KTSNKNLSIVLDFCSGSGTTLKASFLQRFKFGIDNSDLAIKACKNKLETTIKDLFVS
QNPYDFLV2"

complement (3793..4578)

/notes="JHP46-like protein"

/codon_start=1

/transl_table=11

/product="type II restriction enzyme"

/protein_id="AAL37450.1"

/db_xref="GI:17225520"

/translation="WVDYVTKLSIEYANORSYLDLFLQVYPTIPEGLREIDSKWSNV
EYHFKGLALITELLNLDLFPIDSYMAYLRDKXSALERNPTINRICRLYBMGL
NKIFECSPKQTNQIGPMFKWNNKSLGVEPVDLNDFIANENDAILRASDNIME
FAKSHLYHHHGLDFVARFNKKYIIGEAFLTFDFGHNQAFNDIAITISAPNIKAI
KVAILDGLVIESNNQRKLLDTYRNYNIMSALVLRFLYQI"

ORIGIN

Query Match 3.1%; Score 180.8; DB 1; Length 5001;

Best Local Similarity 51.6%; Pred. No. 6e-28;

Matches 472; Conservative 0; Mismatches 427; Indels 15; Gaps 2;

QY	1664	ACCTTCTAGTGAAGCGTTTGACCGCTCCAGGAGGCATCTATGCCGATGATCGCGC	1723
DB	2813	AGCTTGTGTCAGCTTTGATAGCAAAATCGAAATATCAATGCTATAAATTTTCTG	2872
QY	1724	TTTAAAGGGGTGAGGCTATAAGCGTAGTACCGGAGCTTCGAAAGGATCGAGCACTAA	1783
DB	2873	TGCAATAAAAAACGAGATTTTAAGTGTGTCAGAACCAAAAAACAATCTAAAAAATG	2932
QY	1784	TCCCCCTGTTACTCCCTCTTTGGACGATGAGCTTGAGCATGCCAGATTTTCTCGGTG	1843
DB	2933	CTATTTTATTAGAGAGGCTCTAATGATTAAATCTAATAATTAGACGCTTTTCTGTGA	2992
QY	1844	GGGTATCGGGGTACGAGGATCCTTGAACCTGCCAAACGCTCTGGAGCTTCTCCCTTC	1903
DB	2993	GGGTAGCTGGATATTGCGGCTTTAAATTCAGATGCTCTGGACTCTTTTGGC----	3048
QY	1904	TTCAGGCGATCCGACGCTAAACTTCTTCGGGGACACCCGCTCTTTGACCAACATA	1963
DB	3049	--AACTGTCTTAAGGCAATAAATTTTTTCTAGGGTTATTATTATTAGACTCAATC	3106
QY	1964	AGCCCTTGAGCGCTAGCTCGTCAAGCTTCTCCGGGGATAGCGCAATGCCGTCACGGA	2023
DB	3107	AAACCTTCTTTGCCAAGCTCAAGTGTGGCAATATCAGTGGCAATGCGGCCCTTTT	3166
QY	2024	GGGGGAAGTATTCTCTGCCAAGCCCTTCGGGTAGGCGCATCTCTGTTTCTCCAGAGA	2083
DB	3167	GGAGGTAGCATACCTTTAAATGCTTTTAGAACATTGCCCACTTTCACCTTCTCTGGA	3226
QY	2084	TGCAGGGGATTGGTGGTGTACCGTTCCCGGTTCTCGTCTACAAAGGGGAAAGCTACG	2143

DB	3227	TGATTGGAACGGTAGTGTAAACGCCCTTTTATCTTTGTCAATTTTAGGGAATCGTTTTCT	3286
QY	2144	ATCTCCTCTCCGAAATAGGGCTAGCGATTCTGTTCCAAACGTAGTCCCGCGTTTGGAG	2203
DB	3287	AAATCTTGTGGCGTATAGGAATCTTAGGTTCTGTAAAAATGGGATTTTCTCTTAGAG	3346
QY	2204	TAGACGAGGATCATGTCCTTTTCCGATCCGAAAGCCCTTACGGGAAAAGTTTTGGGATTT	2263
DB	3347	TAAATAAAATCATATCTTTTATGTTGCCATAGCCCATCTCTTTAAAAATTTTAGGATTG	3406
QY	2264	GAACGATGGGGGATATGTTAAACGAGTTTTCGCGGCCCAAGACCTCATCAAGGATG	2323
DB	3407	CACTTTATGGGTGATTTTCATTTCTTAAAAATTTTATGTCCAAATATTTTATCTAACATC	3466
QY	2324	AGCTTACCTCGAACCCGATATTTCTCGTCTATGTGAACGAAGATCATGCTAGTCCGCC	2383
DB	3467	ACCTTAAATAATGCCCCTATCTTACAATCTGTATGCACATAGATAGAGCCTTCTCTGAA	3526
QY	2384	ATCAGCTCCCTGAGAAGTATCAAGCGTCCCTCAGGACTCCACAACTGAGGACCATCG	2443
DB	3527	AGCAATTTCTTTAAGCAATCTAGCGGTGCTTTAAAAATTTCTATAAAATCCATCCOACT	3586
QY	2444	AGGGTGTATCGTAGCCCAACTGACCGTTTTCGGGCTGGCTGACGGTAGCAACCGCATCT	2503
DB	3587	ACTTTGTGCTATAGCAATATCGCCATCTTAGAATTCGTAATTTGTTGTTGCTCTACCA	3646
QY	2504	GTTTCATCGCGGCAACAGAAACTGCTGCGCGTTCATAGCGGGTCCATATATAGACC	2563
DB	3647	TTT-----GTGATAGTAAAAATGTTTGTGTCAAATGGAGGTCAATATATATC	3697
QY	2564	AACTGGACCTTCCC 2577	
DB	3698	AAATCTATTTTCC 3711	

RESULT 13

ARI39425/C

LOCUS ARI39425 180 bp DNA linear PAT 16-JUN-2001

DEFINITION Sequence 5 from patent US 6207377.

ACCESSION ARI39425

VERSION ARI39425.1 GI:14481921

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 180)

AUTHORS Wayne, J. and Xu, S.-Y.

TITLE Method for construction of thermus-E. coli shuttle vectors and

identification of two Thermus plasmid replication origins

JOURNAL Patent: US 6207377-A 5 27-MAR-2001;

FEATURES Location/Qualifiers

source 1..180

/organism="unknown"

/mol_type="unassigned DNA"

ORIGIN

QY	4128	CTCTTCAAGAAAGTTTTTTCGTTCTTACCCCTCGGACCTCTTGTTCATCTGAGAGCCGA	4187
DB	180	CTCTTCAAGAAAGTTTTTTCGTTCTTACCCCTCGGACCTCTTGTTCATCTGAGAGCCGA	121
QY	4188	GGCGTTACCCCTAGTCTTGGGGGTGATCCGGGCAACCGCCTCGGTTTCGCCCTTTTATG	4247
DB	120	GGCGTTACCCCTAGTCTTGGGGGTGATCCGGGCAACCGCCTCGGTTTCGCCCTTTTATG	61
QY	4248	GGTCCAAAATAACCGTCAGCCCGGCTGGCAATCCCCCTCTCTAAAAGCGGTTATAG	4307
DB	60	GGTCCAAAATAACCGTCAGCCCGGCTGGCAATCCCCCTCTCTAAAAGCGGTTATAG	1


```

RESULT 14
AF328910      3184 bp      DNA      linear      BCT 12-NOV-2002
LOCUS
DEFINITION
Helicobacter pylori strain CR4 type II DNA modification enzyme,
type II DNA modification enzyme, and type II restriction enzyme
genes, complete cds.
ACCESSION
AF328910
VERSION
AF328910.1 GI:17225500
SOURCE
Helicobacter pylori
ORGANISM
Helicobacter pylori
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
Helicobacteraceae; Helicobacter.
1 (bases 1 to 3184)
Chanto, G., Occhialini, A., Gras, N., Alm, R. A., Megraud, F. and
Marais, A.
Identification of strain-specific genes located outside the
plasticity zone in nine clinical isolates of Helicobacter pylori
Microbiology 148 (Pt 11), 3671-3680 (2002)
12427957
2 (bases 1 to 3184)
Chanto, G., Occhialini, A., Gras, N., Megraud, F. and Marais, A.
Direct Submission
Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
Location/Qualifiers
1. 3184
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="CR4"
/specific_host="Homo sapiens"
/db_xref="taxon:210"
/map="region F between JHP43/HP50 and JHP47/HP55"
/note="isolated from patient with gastric carcinoma"
complement(1..561)
/function="methyltransferase"
/note="JHP44-like protein"
/codon_start=1
/transl_table=11
/product="type II DNA modification enzyme"
/protein_id="AAL37436.1"
/db_xref="GI:17225501"
/translation="MLVSRFLNDIDPFLGVLSRFOIKNGCIYGVCSVKASKFTSS
YEESKAOVLNIIISLKHQIQSNQESVTKVGTFFVILENDLNLDENAFYKLLNLL
INDPFPFLVTPSNGTSPHLSRITPREARIQSFSNVIYFGNKTISVCKQIGN
AVPPLALALGKALIKSTRNDINLSR"
complement(932..1963)
/function="methyltransferase"
/note="JHP45-like protein"
/codon_start=1
/transl_table=11
/product="type II DNA modification enzyme"
/protein_id="AAL37437.1"
/db_xref="GI:17225502"
/translation="MQLNLIQENAIALLFLNDKNLKGKIDLIYIDPPFANNHFTI
TNGRATTISNKGNDIAYSDKVGMDFEFLKORLVLLKELLISBQGSIIYVHTDCKIGH
YVQMLDIFGQFNPRNITRIKCNPKFKRIGYNIKDMILFYSGKNLFPNPKPI
YTPQDLRFPKLDKRRYTPVPHAGEVESGECSEKAFKGMPLPKGRHWRTDIATL
EWDKEGLEIYSNNPKRIYALQAGRVQDIWEFKDPQFPSPYFEKNAQLDLII
KTSNKSISIVLDFCGSGTLLKSAFLQRFKIFGIDNSGLAIQACKNKLEITIKDLFVS
QNFDYDLFV"
complement(1973..2758)
/note="JHP46-like protein"
/codon_start=1
/transl_table=11
/product="type II restriction enzyme"
/protein_id="AAL37438.1"
/db_xref="GI:17225503"
/translation="MVNWTKLSEIYANQSRSYLDLDFQVYPTIPGRLREIDSKIWSNV
EYHFKKNLALITELLNLDLPIKOSVYAVLKROKSALEARNPTINRICGRLYEMGL
NKIFKCSBPKEINQIGPMFKDWINKSLGVEPVDLNDFTANENDALIRASDNIMAE
FARSHLNYHHKGLDFVARPNKKYIIIGBAKFLTDFGCHQNAQFNDALSTIAPNIKA

```

ORIGIN

```

Query Match      3.1%; Score 179.2; DB 1; Length 3184;
Best Local Similarity 51.5%; Pred. No. 1.4e-27;
Matches 471; Conservative 0; Mismatches 428; Indels 15; Gaps 2;
QY 1664 ACCCTTCTAGTGAAGGCTTTGACCCGCTCCAGGAGGATCTATCCGATGATCGCGC 1723
DB 993 AGCTTGTGTTTTCGAAGCTTGGATAGCTAAACGGHAATTATCAATGCTTATAAATTTG 1052
QY 1724 TTTAAGAGGGGTGAGGCTATAGCGCTAGTACCGGAGCCTCGGAGGATCGAGCACTAAA 1783
DB 1053 TGCATATAAAATGCAGATTTTAAAGTTGTTCCAGAACACAAAACAATCTAAAACAATG 1112
QY 1784 TCCCCCTCGTTACTCCCTGTTTGGAGAGTACGCTTGAGCATGTCAGATTTTCTCGGTG 1843
DB 1113 CTATCTTTATTAGAAGAGGTTTTTAATGATTAATACTAATATGAGCGTTTTTCTGTG 1172
QY 1844 GGGTATCGGGGTGCGGAGGATCCTTGAACTGCCAAACGCTCTGGAGCTTCTCCCTTTC 1903
DB 1173 GGGTACTTGGATATTGTGGGCTTTTAATTCAGATGTCTTGGACTCTTTGGCAGCT 1232
QY 1904 TTGAGCGCATCCGAGCGTAAACTTTCTTCCCGGGCACCCCGTTCTTTTGACAGACAATA 1963
DB 1233 T-----GTTCTAAGGCATAAAATTTTTTTCTAGGGTTATTATGTTAGATACTCAATC 1286
QY 1964 AGCCTTGAGCGCTAGCTCGTCAAGCTTCTCGGGGGATAGCGCAATGCCGTCCAGGA 2023
DB 1287 AAACCTTCTTTATCCCAAGCTCAAGTGGCAATATCAGTGGCAATGCCGCTTTT 1346
QY 2024 GGGGGAAGTATTCCTCGCAAGCGCTTCGCGTAGGCGCAATCTTGTTTCTCCAGAGCA 2083
DB 1347 GAGGTAGTCATACCTTTTAAATGCTTTAGAGCATTTGCCACTTTCTACTTCTCTCGGAGCA 1406
QY 2084 TGAGGGGATTTGGTGTGTACCGTTCCCGGTTCTCGTCTACAAAGGGGAAAGCTTAGCG 2143
DB 1407 TGATTGGAAACGCTAGTGAACGCTTTTATCTCTGTCAAATTTTGGGAATCGTTTTTCT 1466
QY 2144 ATCTCTCTTCGGAATAGGGGCTAGCGGATTCGTTCCAAACGATGTCGCGCGTTTGGAG 2203
DB 1467 AAATCTTGTGGCGTATPAGGAATCTTAGGCTCGTTAAAAATGGGATTTTTCCTTTAGAA 1526
QY 2204 TAGCAGGAGTATGTCCTTTTTCGATCCGAGCGCTTACGAGGAAAGTTTGGGATTT 2263
DB 1527 TAAATATAAATCATATCTTTTATGTTACCATAGCTATCCTTTTAAAAATTTTGGATTG 1586
QY 2264 GAGCGATCGGGCGATATGTTTACGAAGTTTTCGCGGCAAGACCTCATCAAGGATG 2323
DB 1587 CATTATAGCTGTGATTTCAITTTTAAATTTTGTATGCCAAATATTTCATCTACATC 1646
QY 2324 AGCTTCACCTCGAACCCGTTATTTCTCGTCTATGTGAACGAAGATCAGTCTGAGTCCGCC 2383
DB 1647 ACCTTAACATAATGCCCTATCTTTACAATCTGTATGCACATAGATAGAGCCTTCTTGAA 1706
QY 2384 ATAGCTCCCTGAGAGTATCAAGCGCTCCCTCAGGAATCCCAAACTGAGAGCAGCATCG 2443
DB 1707 AGCAATTTCTTAAAGCAATAGTAGCGGTGCTTTTAAAAATTTCTATAAATCCATCCACT 1766
QY 2444 AGGTGTCTATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACCGGATCT 2503
DB 1767 ACTTTGTCGTATAAGCAATATCGGCATTCTTAGAATTTGTAATTTGTTGTTCTGCCA 1826
QY 2504 GTTTCATCGCGGCAACGAGAAACTGCTGGCCGGTTCATTAAGCGGGTCAATATAGACC 2563
DB 1827 TTT-----GTGATAGTAAAAATGATTTGTTGTAACAAATGAGGGTCAATATATATC 1877
QY 2564 AACTGGACCTTCCC 2577
DB 1878 AAATCTATTTTCC 1891

```

RESULT 15

AF328916 4931 bp DNA linear BCT 12-NOV-2002
LOCUS
DEFINITION Helicobacter pylori strain CR35 cytosine-specific DNA
methyltransferase, putative HP0052-like protein, type II DNA
modification enzyme, and type II restriction enzyme genes, complete
cds.
ACCESSION AF328916
VERSION AF328916.1 GI:17225521
KEYWORDS
SOURCE
ORGANISM Helicobacter pylori
Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
Helicobacteraceae; Helicobacter.
REFERENCE 1 (bases 1 to 4931)
AUTHORS Chanto, G., Occhialini, A., Gras, N., Alm, R.A., Megraud, F. and
Marais, A.
TITLE Identification of strain-specific genes located outside the
plasticity zone in nine clinical isolates of Helicobacter pylori
JOURNAL Microbiology 148 (Pt 11), 3671-3680 (2002)
PUBMED 12427957
REFERENCE 2 (bases 1 to 4931)
AUTHORS Chanto, G., Occhialini, A., Gras, N., Megraud, F. and Marais, A.
DIRECT SUBMISSION
TITLE Submitted (12-DEC-2000) Laboratoire de Bacteriologie, Universite
JOURNAL Bordeaux 2, 146 rue Leo Saignat, Bordeaux 33076, France
FEATURES
source
1. 4931
/organism="Helicobacter pylori"
/mol_type="genomic DNA"
/strain="CR35"
/specific_host="Homo sapiens"
/db_xref="taxon:210"
/map="region F between JHP43/HP50 and JHP47/HP55"
/notes="isolated from patient with chronic gastritis"
complement(1..1086)
/notes="HP0051-like protein"
/codon_start=1
/transl_table=11
/product="cytosine-specific DNA methyltransferase"
/protein_id="AAL37451.1"
/db_xref="GI:17225522"
/translation="MNYKLDLFCGAGGFSAGLYLKEPDLIGLDDCKOALITPENN
HKAIGVCGDITQTEIKERVIELAKLEINLIIGPCGFSNKNKNGILKDPENFLF
LEWIEVKAKEIFEIENKVLISCAGKYFLEEIKERLALGVLGSLQILNADYGV
PQNRKAFVAGSRFSDFNLELPSGVNVQDAISDLAYLCNEGAFESDYNIQSS
YQALMKRFPKYNHOATHSQAALKLKNKEQKCLPKNHLHGKQFSTWGRLN
WKISPTIDTFPSNGTNSHPLEHRSITPREAARIQSPDDYIFYGNKTSVCKQIG
NAVPPLLALGKALIKSARNDTNLSR"
complement(1083..2309)
/codon_start=1
/transl_table=11
/product="putative HP0052-like protein"
/protein_id="AAL37452.1"
/db_xref="GI:17225523"
/translation="MLVSRFLNDIDFPNLGLVLSRFQIKNGCIYGVCSYKASKFVPD
YEKSAVDALNTSVHPWOSNOESVTIKIGTFVFLDKDLHLDENAFYKLLNLI
INDPNRSHWTKQKRLPSGFFSRGSDIQORNFSLDYEFHSPLEFNKPHYLIDH
ENIPEALNFNRELOPEHAQGINORNAQFRIYLNWYHLHGLNPNYKARAEHVFT
TILYDIYHKLSPYRTIKHNSPTEHAFYLNKYQQDDLDKSEKLEQLQGIQKS
EFKRSKILNFIYRSTPNVSCCGDYDIKERSFISLPYKIQTNPSNYTEIHVV
SLGNKELDVLENLAKLCPCTCHRALKGSSEEEFQRLIRNLNHNKDNLEFAQLRFE
TDFFLTLDIKHESLK"
complement(2676..3707)
/function="methyltransferase"
/notes="JHP45-like protein"
/codon_start=1
/transl_table=11
/product="type II DNA modification enzyme"
/protein_id="AAL37453.1"
/db_xref="GI:17225524"
/translation="MQNLLIQENAIALLFLNDKNLKGKIDLIIVIPDPATNHHFTI
TNGRATTISNSKNGIAYSDKVGMDFIEFLKQRLVLLKELLSQGSIIYVHTDCKIGH
YKVMLEDFIGIQNFENITRIKCNPKNFKIRIGYIKMDILFYSGKNPFFNPKIP

VTPODLKRPRPKIDKRRYTVTPIHAPGEVSECSKAFKGM.LPKGRHWRDITL
ERWKEGLIIEYNNNNPRKKIYALEQAGKRVQDIWEKDFQYPSYFTEKNAQLDLII
KTSNKGSIVLDDCFGSGTTLKSAFLQRRFIGIDNSNLAIQACKNKLETITTKDLFVS
QNFDYFLVF"
complement(3717..4502)
/notes="JHP46-like protein"
/codon_start=1
/transl_table=11
/product="type II restriction enzyme"
/protein_id="AAL37454.1"
/db_xref="GI:17225525"
/translation="MVDYWTKLISYANQSRYSYLDLQVYVYTIPEGLREIDSINNV
EYHFKRDNALILTELLNLDLFFPKDSYMAYLKRDKSALERNPTNIRICGRYEMGL
NKIEKCESEPKETNROIGPMFKDWNNKSLGVEPDLNDFIANENAILRASNNIMAE
PAKSHLVNHHKGLDFVARNKYVILGEAKFLTDFGHNQAFNDALISIEAPNIKAI
KVALDGVLYIESNNKRLDITYRNYNIMSALVLRFLYQI"
ORIGIN
Query Match 3.1%; Score 179.2; DB 1; Length 4931;
Best Local Similarity 51.5%; Pred. No. 1.3e-27;
Matches 471; Conservative 0; Mismatches 428; Indels 15; Gaps 2;
QY 1664 ACCCTTCTAGTAGAGGCTTTGACCGCCTCCGAGGAGCATCTATGCCGATGGATCGCCG 1723
DB 2737 AGCTTGTTTTGCAAGCTTGATAGCTAAATGGATATATCAATGCTATAAATTTTCGT 2796
QY 1724 TTTAAGAGGGTGAGGCTATAAGCGTAGTACCGGAGCTCGAAGGATCGAGCACTAAA 1783
DB 2797 TGCATATAAACCGCAGATTTTAGGTTGTTCCAGAACCAAAAACAATCTAAACAATG 2856
QY 1784 TCCCCCTCGTTACTCCTGTTTGGACGATGAGCTTAGCATGTCCAGATTTTTTCTCGGTG 1843
DB 2857 CTACCTTTATTAGAAGAGGTTTTAAATGATTAAGTCTAAATAATTAGCGGTTTTTCTGTG 2916
QY 1844 GGGTATCGCGGTACGAGGATCCTTGAACCTGCAACGCTCGTGGAGCTTCTTCCCTTC 1903
DB 2917 GGGTAGCTTGATATTGTTGGGCTTTTAAATCCAGATGCTTTGGACTCTTTTGCAGCT 2976
QY 1904 TTCAGGCGATCCCGAGCGTAAACTTTTCCCGGGGACCCCGTTCTTTTGACCAACAATA 1963
DB 2977 T-----GTTCTAAGGCATAAATTTTTTTTCTAGGGTTATTATTGTTAGATACTCAATC 3030
QY 1964 AGCCTTTCAGCGCTAGCTCGTCAAGCTTCTCGGGGGATAGCGCAATCGCTCCGAGCA 2023
DB 3031 AAACCTTCTTTATCCCAACGCTCAAGTGTGGCAATATCAGTGGCGCAATGCGCCCTTTT 3090
QY 2024 GGGGGAAGTATTCCTCGCCAAAGGCTTCCGCTAGGGCCATCTTGGTTTCTCCAGGAGCA 2083
DB 3091 GGAGGTAGCATACCTTTAAATGCTTTTAGAGCATTCGCCACTTTCTACTTCTCCTGGAGCA 3150
QY 2084 TGCAGGGGATGTTGGTGTGTACCGTTCCCGTTCTCTACAAAGGGGAAAAGCCTAGCG 2143
DB 3151 TGTATTGGAAACGGTAGTGTAAACGCTTTTATCCTGTCATTTTAGGGAACTCGTTTTCT 3210
QY 2144 ATCTCCTCTTCCGAATAGGGGCTAGCCGATTCGTTCCAAACGTAGTCCCGCGTTTGGAG 2203
DB 3211 AAATCTTGTGGCTATAAGGAATCTTAGGCTCGTTAAATAATGGATTTTTTCTTTAGAA 3270
QY 2204 TAGACGAGGATCATGCTCTTTTCGATCGAAGGCTTACGGGAAAAGTTTGGGATTT 2263
DB 3271 TAAATAAATCATATATCTTTTATGTTACCATAGCCCTATCCTTTTAAATTTTTTAGGATTTG 3330
QY 2264 GAAGCGATGGGCGGATATGTTTAAAGAAATTTCCGCGCCGCAAGACCTCATCAAGGATG 2323
DB 3331 CACTTTATAGCTGTGATTCATTTCTAAATTTTGTATGCCAATAATTTCACTACATC 3390
QY 2324 AGCTTCACCTCGAACCCTGATTTTCTCGTCTATGTGAACGAAGATCAGTCTGAGTCCGCC 2383
DB 3391 ACCTTAACATAATGCTCCCTATCTTACAATCTGTATGCACATAGATAGAGCCTTTGTTCTGAA 3450
QY 2384 ATCAGCTCCCTGAGAGTATCAAGGCTCCCTCAGGAACCTCCACAACTGAGGACCATCG 2443
DB 3451 AGCAATCTTTTAAAGCAATACCAAGCGCTGTTTTTAAATAATTTTATAAATCCATACCCT 3510

Qy	2444	AGGCTGTCATCGTAGCCCAACTGACCGTTTTTGGGCTGGCTGACGGTAGCAACGGCATCT	2503
Db	3511	ACTTTGTCGGCTATAAGCAATATCGCCATCTTAGAATTGCTAATTGGTTGCTCTGCCA	3570
Qy	2504	GTTCATCGCGCCCAACGAGAACTGCTGCGGTTCCATAAGCGGGTCAATATAGACC	2563
Db	3571	TTT-----GTGATAGTAAATGATTTGTAGCAATGGAGGTCAATATATC	3621
Qy	2564	AACTGGACCTTCCC	2577
Db	3622	AAATCTATTTTCC	3635

Search completed: March 18, 2004, 02:39:47
Job time : 14877.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 20:20:31 ; Search time 1429.28 Seconds
(without alignments)
17384.739 Million cell updates/sec

Title: US-09-664-186-6
Perfect score: 5849
Sequence: 1 tctagaagtcagggtggac.....tgtagagaccagaagacc 5849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5849	100.0	5849	5 AAD04668	Aad04668 Thermus p
2	1242	21.2	1242	2 AAX01848	Aax01848 Thermus s
3	1026	17.5	1026	5 AAD04666	Aad04666 Thermus r
4	999	17.1	999	2 AAX01849	Aax01849 Thermus s
5	185.6	3.2	1032	2 AAT67772	Aat67772 H. pylori
6	185.6	3.2	1032	2 AAT77452	Aat77452 H. pylori
7	180	3.1	180	5 AAD04667	Aad04667 Thermus p
8	59	1.0	2000	7 ADA71938	Ada71938 Rice gene
9	57.6	1.0	2000	7 ADA71938	Ada71938 Rice gene
10	56.4	1.0	2030	7 ADA70342	Ada70342 Rice gene
11	56.4	1.0	2031	9 ADC08176	Adc08176 Rice DNA
12	55.4	0.9	2028	8 ADA48615	Ada48615 Rice gene
13	55.4	0.9	2028	9 ADC08122	Adc08122 Rice DNA
14	54.8	0.9	2028	7 ADA70341	Ada70341 Rice gene
15	54.8	0.9	2060	7 ADA71932	Ada71932 Rice gene
16	54.4	0.9	615	2 AAQ37036	AAQ37036 SOD-r gen
17	53.2	0.9	1755	8 ADA70624	Ada70624 Rice gene
18	53.2	0.9	1965	8 ADA48223	Ada48223 Rice gene
19	53.2	0.9	1965	9 ADC08174	Adc08174 Rice DNA
20	52.4	0.9	113193	7 AAD54645	Aad54645 Streptomy
21	51.6	0.9	1815	5 AAD09400	Aad09400 Zea mays
22	50.2	0.9	509	8 AC123248	AC123248 DNA clone
23	49.6	0.8	1674	7 ACA38419	ACA38419 Prokaryot

C	24	49.6	0.8	1677	7	ACA40670	ACA40670 Prokaryot
	25	49.6	0.8	110000	4	AAI99682_27	Continuation (28 o
	26	49.6	0.8	110000	4	AAI99683_27	Continuation (28 o
	27	49.4	0.8	349800	6	ABQ81846	Abq81846 Bifidobac
C	28	48.8	0.8	1236	9	ADC36241	Adc36241 Weed cont
C	29	48.8	0.8	1454	9	ADC36242	Adc36242 Weed cont
C	30	48.6	0.8	21034	2	AAV62154	Aav62154 HSV-2 str
	31	48.6	0.8	117213	2	AAV62176	Aav62176 HSV-2 str
	32	48.6	0.8	154746	6	AAD25519	Aad25519 Human har
C	33	47.8	0.8	963	7	ACA24030	ACA24030 Prokaryot
	34	47.4	0.8	10732	3	AAAI0594	AAAI0594 Gene enco
C	35	47.2	0.8	1194	7	ADA69660	Ada69660 Rice gene
C	36	47	0.8	6094	3	AAA71428	Aaa71428 Expressio
C	37	46.8	0.8	1605	7	ABT31955	Abt31955 High-affi
C	38	46.6	0.8	924	6	ABK49877	Abk49877 Environme
C	39	46.6	0.8	7659	7	ACD13414	Acdi13414 Human DNA
C	40	46.6	0.8	10065	7	AAL54465	Aal54465 Human cad
C	41	46.6	0.8	10065	7	ACA63037	ACA63037 Human cad
C	42	46.6	0.8	10455	7	AAL54468	Aal54468 Human cad
C	43	46.6	0.8	10455	8	ACA63040	ACA63040 Human cad
C	44	46.6	0.8	11073	7	AAL54466	Aal54466 Human cad
C	45	46.6	0.8	11073	8	ACA63038	ACA63038 Human cad

ALIGNMENTS

RESULT 1
AAD04668
ID AAD04668 standard; DNA; 5849 BP.
XX
AC AAD04668;
XX
DT 04-JUL-2001 (first entry)
XX
DE Thermus plasmid pTsp45S DNA sequence.
XX
XX Replication protein; Rept; partition protein; Para; pTsp45S plasmid;
XX kanamycin-resistance gene; thermophilic transformation; Ori;
XX replication origin; ds.
XX
OS Thermus sp.
XX
PN US6207377-B1.
XX
PD 27-MAR-2001.
XX
PF 14-AUG-1998; 98US-00134246.
XX
PR 14-AUG-1998; 98US-00134246.
XX
XX (NEWE) NEW ENGLAND BIOLABS INC.
XX
Wayne J, Xu S;
WPI; 2001-298939/31.
XX
XX Cloning Thermus species (Ts) plasmid genes comprises transforming
XX Escherichia coli with cloned recombinant plasmid containing Ts and E.coli
XX origins of replication, isolating cloned recombinant plasmid from E.coli
XX and transforming Ts cell.
XX
XX Example 1; Fig 3; 32pp; English.
XX
XX The present sequence is Thermus plasmid pTsp45S DNA. The open reading
XX frame of pTsp45S plasmid is the replication protein. Rept which is needed
XX for thermophilic plasmid replication. The invention relates to Thermus
XX sp. replication protein Rept, partition protein Para and their
XX corresponding DNA molecules which relates to recombinant DNA molecules
XX encoding plasmid DNA replication origins in Thermus, as well as to
XX shuttle vectors which contain the same. The invention also relates to
XX method useful for cloning Thermus sp. plasmid genes which comprises
XX inserting plasmid DNA comprising a Thermus sp. origin of replication

CC (Ori) into a recombinant plasmid comprising a thermostable kanamycin-
CC resistance gene and an Escherichia coli Ori, to produce a cloned
CC recombinant plasmid. This cloned recombinant plasmid is transformed with
CC an E. coli. host cell, and E. coli. host cell cultured for the expression
CC of cloned recombinant plasmid. The cloned recombinant plasmid isolated
CC from E. coli host cell is then transformed with Thermus sp. host cell and
CC Thermus sp. host cell is cultured. Thus Thermus sp. plasmid genes are
CC cloned. These plasmid DNAs are used for thermophilic transformation
XX
SQ Sequence 5849 BP; 1377 A; 1501 C; 1670 G; 1301 T; 0 U; 0 Other;
Query Match 100.0%; Score 5849; DB 5; Length 5849;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5849; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAGAGGTGACGGGTGGCAAGGAAACACCAATAGCCCTCCAGAGAGATGGACGAGT 60
DB 1 TCTAGAGGTGACGGGTGGCAAGGAAACACCAATAGCCCTCCAGAGAGATGGACGAGT 60
QY 61 TGGTGTCCGAAAGTGGCCATCCGGGGCGCTTTCACAACTATTTCCAGCGGTGGCA 120
DB 61 TGGTGTCCGAAAGTGGCCATCCGGGGCGCTTTCACAACTATTTCCAGCGGTGGCA 120
QY 121 CCGGCATTGGCCACAGAGTACGAGCTTGTGGAGTAGACGGCCACAAAGGGGTGCTCTCA 180
DB 121 CCGGCATTGGCCACAGAGTACGAGCTTGTGGAGTAGACGGCCACAAAGGGGTGCTCTCA 180
QY 181 AACTCTTTCTAGTCGCTTGGACGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
DB 181 AACTCTTTCTAGTCGCTTGGACGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 240
QY 241 CCTCTTCCCTCCTCTCTGGGGCGCTTAGCGCGCTTAAACCTGAGACGGGCTCAAGTT 300
DB 241 CCTCTTCCCTCCTCTCTGGGGCGCTTAGCGCGCTTAAACCTGAGACGGGCTCAAGTT 300
QY 301 TAGGGATTTCGCTTTCGGGGTAAGAAATCCGGCGCTCAGGGGATCGCGGATGGCCCTTA 360
DB 301 TAGGGATTTCGCTTTCGGGGTAAGAAATCCGGCGCTCAGGGGATCGCGGATGGCCCTTA 360
QY 361 TCCTGCGGCTCCCTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
DB 361 TCCTGCGGCTCCCTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
QY 421 CTTCTCTAACAGAGACACAAACCTCTAAACCTATCAATCCACCGATTCAGATATA 480
DB 421 CTTCTCTAACAGAGACACAAACCTCTAAACCTATCAATCCACCGATTCAGATATA 480
QY 481 CCAATAATGGCACAAGTTTGGAGAGTGGTCAACAAAGAGGCTTTCGGTCAGGTT 540
DB 481 CCAATAATGGCACAAGTTTGGAGAGTGGTCAACAAAGAGGCTTTCGGTCAGGTT 540
QY 541 ATGGTCAAGTGGGGCGGTCAAGGCGGCTTAAAGTTGGTAAAGCGGGAGAGCAAA 600
DB 541 ATGGTCAAGTGGGGCGGTCAAGGCGGCTTAAAGTTGGTAAAGCGGGAGAGCAAA 600
QY 601 CCGGGTGTATCAATGCAACAGATGCGAGTGGAAACGTTGGACACAGAGAGGCTTGA 660
DB 601 CCGGGTGTATCAATGCAACAGATGCGAGTGGAAACGTTGGACACAGAGAGGCTTGA 660
QY 661 GCTTCTGGAGAGGGGTATTTGGATAAATCTGACAGGCTTATAAGGGGAAAGTGGCTC 720
DB 661 GCTTCTGGAGAGGGGTATTTGGATAAATCTGACAGGCTTATAAGGGGAAAGTGGCTC 720
QY 721 TTCAGGTTCAGTACAGAGAGGTAGAGGAAAACTTCGCGAGGCTCAAGGCAATACGA 780
DB 721 TTCAGGTTCAGTACAGAGAGGTAGAGGAAAACTTCGCGAGGCTCAAGGCAATACGA 780
QY 781 GGGAGGAGGATAGTCCGGGCGCAACAACTCGTGGAGCGGCTCAATATGCCAG 840
DB 781 GGGAGGAGGATAGTCCGGGCGCAACAACTCGTGGAGCGGCTCAATATGCCAG 840
QY 841 AAAAAAGTTCGAGCGGTCCCTCTCAATCACCCCTACCTGCTTTGGTCTACTACCTGGT 900
DB 841 AAAAAAGTTCGAGCGGTCCCTCTCAATCACCCCTACCTGCTTTGGTCTACTACCTGGT 900

DB 841 AAAAAAGTTCGAGCGGTCCCTCTCAATCACCCCTACCTGCTTTGGTCTACTACCTGGT 900
QY 901 TTCGAAAAAGCAGAAAAAGCGAAACAGGCGCTTTGAGAGGCATTTGAGAGGTTGCTTC 960
DB 901 TTCGAAAAAGCAGAAAAAGCGAAACAGGCGCTTTGAGAGGCATTTGAGAGGTTGCTTC 960
QY 961 AAAGCACCAGAAACCATTCGCGTCTTGGCAGAGAAAGCGCAAGAGAGCGGTAGAGC 1020
DB 961 AAAGCACCAGAAACCATTCGCGTCTTGGCAGAGAAAGCGCAAGAGAGCGGTAGAGC 1020
QY 1021 CTTGATCCAAAGGCTCAAGGAGCTCCCAATAAATCGGCAGATAGGCGCGGTGTTCAA 1080
DB 1021 CTTGATCCAAAGGCTCAAGGAGCTCCCAATAAATCGGCAGATAGGCGCGGTGTTCAA 1080
QY 1081 AAGTGTGTACAAAGAGAGCTAAAGGGGAAATAGAGAGAGGCTTCCAGGCGCTTACCAA 1140
DB 1081 AAGTGTGTACAAAGAGAGCTAAAGGGGAAATAGAGAGAGGCTTCCAGGCGCTTACCAA 1140
QY 1141 ACCAAAGATTGTGTAGTATCCCTGAAAGAAAGTAAACCGGAGCAAGCACCCTTATTGC 1200
DB 1141 ACCAAAGATTGTGTAGTATCCCTGAAAGAAAGTAAACCGGAGCAAGCACCCTTATTGC 1200
QY 1201 GGAGAGAGAGCGGCGCATCATATACACGGGATCGGATGAAGCTTTGAAGATGCGC 1260
DB 1201 GGAGAGAGAGCGGCGCATCATATACACGGGATCGGATGAAGCTTTGAAGATGCGC 1260
QY 1261 CAAGGAAAACTGGGCGCTTGGCGAGGAGCAGAACTAGGCACCAAGGGGTAGATTCTA 1320
DB 1261 CAAGGAAAACTGGGCGCTTGGCGAGGAGCAGAACTAGGCACCAAGGGGTAGATTCTA 1320
QY 1321 CGTGGTCATCCGGCTAGCCCTCAAGAGACATGGCACCTAAACAGGAGAGTGAAGTTTCA 1380
DB 1321 CGTGGTCATCCGGCTAGCCCTCAAGAGACATGGCACCTAAACAGGAGAGTGAAGTTTCA 1380
QY 1381 ATCCGACTTTGGCGGAAACCAAGACATAGTGGTGGAGCGAATCGCTGTGTGAGCAA 1440
DB 1381 ATCCGACTTTGGCGGAAACCAAGACATAGTGGTGGAGCGAATCGCTGTGTGAGCAA 1440
QY 1441 GGACCTTGAGAGAGGACACATAGGAATAGTGGTGGAGCGAATCGCTGTGTGAGCAA 1500
DB 1441 GGACCTTGAGAGAGGACACATAGGAATAGTGGTGGAGCGAATCGCTGTGTGAGCAA 1500
QY 1501 GTTTCGTGGTGGCGGACTGGGGAAAGAAACGATCGTTPACATCCGTACTCCTCTTCC 1560
DB 1501 GTTTCGTGGTGGCGGACTGGGGAAAGAAACGATCGTTPACATCCGTACTCCTCTTCC 1560
QY 1561 AGACCTGATAGCGAGCTCTACCAAGGGTGAAGAGCCCTGGGCTCTAGAGGCGGA 1620
DB 1561 AGACCTGATAGCGAGCTCTACCAAGGGTGAAGAGCCCTGGGCTCTAGAGGCGGA 1620
QY 1621 CACAATCTCAAACTGTGTGTAGCCCTGGGAAATCTCTAAACCCCTCTAGTGAAGGC 1680
DB 1621 CACAATCTCAAACTGTGTGTAGCCCTGGGAAATCTCTAAACCCCTCTAGTGAAGGC 1680
QY 1681 TTTGACCGCTCTCCAGGAGCATCTATGCGATGGATCCCGCTTTAAGAGGGGTGAGGC 1740
DB 1681 TTTGACCGCTCTCCAGGAGCATCTATGCGATGGATCCCGCTTTAAGAGGGGTGAGGC 1740
QY 1741 TATAAGCGTATGATACCGGAGCTCGGAAGGATCGAGCACTAAATCCCTCTCGTTACTCC 1800
DB 1741 TATAAGCGTATGATACCGGAGCTCGGAAGGATCGAGCACTAAATCCCTCTCGTTACTCC 1800
QY 1801 TGGTTGACGATAGCTTGAAGATGTCAGATTTTTTCTCGTGGGGTATCGGGGTACGG 1860
DB 1801 TGGTTGACGATAGCTTGAAGATGTCAGATTTTTTCTCGTGGGGTATCGGGGTACGG 1860
QY 1861 AGGATCCTTCAAAGTCCCAAGCGCTCGGAGCTTCTTCCCTCTCTTCCAGCGCATCCCGAGC 1920
DB 1861 AGGATCCTTCAAAGTCCCAAGCGCTCGGAGCTTCTTCCCTCTCTTCCAGCGCATCCCGAGC 1920
QY 1921 GTAAATCTTCTCCGGGCAACCCCGTCTTTGACACAGAAATAAGCCCTTGAGCGTCTAG 1980
DB 1921 GTAAATCTTCTCCGGGCAACCCCGTCTTTGACACAGAAATAAGCCCTTGAGCGTCTAG 1980

QY 1981 CTCGTAAGCTTCTCCGGGGATAGCCCAATGCCGTCAGAGGGGAAAGTATTCCTCG 2040
Db 1981 CTCGTAAGCTTCTCCGGGGATAGCCCAATGCCGTCAGAGGGGAAAGTATTCCTCG 2040
QY 2041 CCAAGGCTTCGGTAGGGCCATCTTGGTTTCTCCAGGAGCATGAGGGGATTTGGTGT 2100
Db 2041 CCAAGGCTTCGGTAGGGCCATCTTGGTTTCTCCAGGAGCATGAGGGGATTTGGTGT 2100
QY 2101 GTACCGTTCCCGGTTCTCGTCTACAAAGGGGAAAGCCTAGCGATCTCTTCCGAATA 2160
Db 2101 GTACCGTTCCCGGTTCTCGTCTACAAAGGGGAAAGCCTAGCGATCTCTTCCGAATA 2160
QY 2161 GGGGTACCGGATTCGTTCCAAAAGCTAGTCCCGGTTTGGAGTAGAGGATCATGTC 2220
Db 2161 GGGGTACCGGATTCGTTCCAAAAGCTAGTCCCGGTTTGGAGTAGAGGATCATGTC 2220
QY 2221 CTTTGGCATCCGAAGGCTTACGGGAAAGTTTTTGGGATTTGAAGCGATGCGGGCGAT 2280
Db 2221 CTTTGGCATCCGAAGGCTTACGGGAAAGTTTTTGGGATTTGAAGCGATGCGGGCGAT 2280
QY 2281 ATGTTAACGAAGTTTCGCGGCCAAAGACCTCATCAAGGATGAGCTTCACTCGAACC 2340
Db 2281 ATGTTAACGAAGTTTCGCGGCCAAAGACCTCATCAAGGATGAGCTTCACTCGAACC 2340
QY 2341 GTATTTCTCGTATGTGAAGAGTCACTCTGAGTCCGCCATCAGCTCCCTGAGAG 2400
Db 2341 GTATTTCTCGTATGTGAAGAGTCACTCTGAGTCCGCCATCAGCTCCCTGAGAG 2400
QY 2401 TATCAAGGCTCCCTCAGGAACTCCACAACTGAGGACCATCGAGGGTGTCTATGAGCC 2460
Db 2401 TATCAAGGCTCCCTCAGGAACTCCACAACTGAGGACCATCGAGGGTGTCTATGAGCC 2460
QY 2461 CAACTGACGTTTGGGCTGGCTGACGGTAGCAAGCGGATCTGTTTCATCGCGCCCAAC 2520
Db 2461 CAACTGACGTTTGGGCTGGCTGACGGTAGCAAGCGGATCTGTTTCATCGCGCCCAAC 2520
QY 2521 GAGAACTGCTGGCGGTTTCCATAGGCGGGTCAATATAGACCAACTGGAACCTTCCCGCG 2580
Db 2521 GAGAACTGCTGGCGGTTTCCATAGGCGGGTCAATATAGACCAACTGGAACCTTCCCGCG 2580
QY 2581 ATACCCACAGGCTCCCGAGCATCCACCGGAGAACCTGACCGTTTTCCCAAAAAGTA 2640
Db 2581 ATACCCACAGGCTCCCGAGCATCCACCGGAGAACCTGACCGTTTTCCCAAAAAGTA 2640
QY 2641 GGTGCCAATAGATCAATCTCAAAAGGGGGGCATTTTCCCTTAGGAGAGGAGGTTTC 2700
Db 2641 GGTGCCAATAGATCAATCTCAAAAGGGGGGCATTTTCCCTTAGGAGAGGAGGTTTC 2700
QY 2701 TTTTCGCAAAACAAAGTTGTGGGTTGGGCTGATCAAGAAATCTCTTCATCGCGTTTTCC 2760
Db 2701 TTTTCGCAAAACAAAGTTGTGGGTTGGGCTGATCAAGAAATCTCTTCATCGCGTTTTCC 2760
QY 2761 GGGGTAGACCAACCTAAAGGGGGAAGTTTCGAGGTTTTTCAGGCTTTCAAGGGGGCTTT 2820
Db 2761 GGGGTAGACCAACCTAAAGGGGGAAGTTTCGAGGTTTTTCAGGCTTTCAAGGGGGCTTT 2820
QY 2821 TCGGGTCAAAACAGGGTAGCTACGGCTCATTTCTCCCTCCCAAGCGCTCTTAAGCAGG 2880
Db 2821 TCGGGTCAAAACAGGGTAGCTACGGCTCATTTCTCCCTCCCAAGCGCTCTTAAGCAGG 2880
QY 2881 ACCTCATCAACCAACCTCAGCACTCCAAAGGAATCCGCAAGGCGGCTTACC 2940
Db 2881 ACCTCATCAACCAACCTCAGCACTCCAAAGGAATCCGCAAGGCGGCTTACC 2940
QY 2941 TTTTCAGCGGCTATCTCCCTGAGTATAGACCTTCGGATCGTCTCAGGGTGACCCCGA 3000
Db 2941 TTTTCAGCGGCTATCTCCCTGAGTATAGACCTTCGGATCGTCTCAGGGTGACCCCGA 3000
QY 3001 AGGATGTCTGCAAGCTCTCGGGGTACGGTACACGGGCTTCACTCATGACCAACCT 3060
Db 3001 AGGATGTCTGCAAGCTCTCGGGGTACGGTACACGGGCTTCACTCATGACCAACCT 3060

QY 3061 TACCCACAGAGGACAAACAATGCGCAACTATGGGCAAGTAGACAACGAGACCAAAAGCTT 3120
Db 3061 TACCCACAGAGGACAAACAATGCGCAACTATGGGCAAGTAGACAACGAGACCAAAAGCTT 3120
QY 3121 GGGCCACTCTCTCAGAGGCTCTTGGGTTTCTCAGTGTAGTCTCCCGTCTGTGTC 3180
Db 3121 GGGCCACTCTCTCAGAGGCTCTTGGGTTTCTCAGTGTAGTCTCCCGTCTGTGTC 3180
QY 3181 AGATGCCATCCGTGTAAATCTCGGTAACCCGCAATAGGAGCGCAATAGATTTGCGG 3240
Db 3181 AGATGCCATCCGTGTAAATCTCGGTAACCCGCAATAGGAGCGCAATAGATTTGCGG 3240
QY 3241 GGGAGTTACCCCTGGCCACAGCCACAGCAAGCTGTGATAGAAAGCTTTGAAACAGCG 3300
Db 3241 GGGAGTTACCCCTGGCCACAGCCACAGCAAGCTGTGATAGAAAGCTTTGAAACAGCG 3300
QY 3301 TCATCTAGGAGATTCGGCAATGTACGTAGCAATGAGGCTGATAGGCGCCGGGAGTTGGAA 3360
Db 3301 TCATCTAGGAGATTCGGCAATGTACGTAGCAATGAGGCTGATAGGCGCCGGGAGTTGGAA 3360
QY 3361 CGCTCACCTCGGGAGGACCGAGATGAGCCAGGTCAACGGCCACGGTCTTGGTGTG 3420
Db 3361 CGCTCACCTCGGGAGGACCGAGATGAGCCAGGTCAACGGCCACGGTCTTGGTGTG 3420
QY 3421 GGCATCAACCTTTTCCCTGAGCCAGAGACAGCACCGTCCAGGCTGAGGCGGATTCOG 3480
Db 3421 GGCATCAACCTTTTCCCTGAGCCAGAGACAGCACCGTCCAGGCTGAGGCGGATTCOG 3480
QY 3481 TGTCTGTGTAGCTTGAACCGCATGTAAGGAGAGACGCGCTTGGCCATGTCTAGGGCG 3540
Db 3481 TGTCTGTGTAGCTTGAACCGCATGTAAGGAGAGACGCGCTTGGCCATGTCTAGGGCG 3540
QY 3541 AGTTCTCCAGGGTAGATGTAGTCCAGGGTAGCTTGGCTTTCCCTGSCCTCAGC 3600
Db 3541 AGTTCTCCAGGGTAGATGTAGTCCAGGGTAGCTTGGCTTTCCCTGSCCTCAGC 3600
QY 3601 CGACCGCCCAAGAGTTCGCGATGCGCCCGCTCCCGTTGACGGTTTGGTGAAGAGC 3660
Db 3601 CGACCGCCCAAGAGTTCGCGATGCGCCCGCTCCCGTTGACGGTTTGGTGAAGAGC 3660
QY 3661 TCGGTGGCCACAGGCGCTTTTCTCAAGACCTTCCAGCGTGGAGCTTCCCGG 3720
Db 3661 TCGGTGGCCACAGGCGCTTTTCTCAAGACCTTCCAGCGTGGAGCTTCCCGG 3720
QY 3721 GTACCCCGCAGGTTGAGGGCCAACTCTCCAGGGGACCATGAAGACGACCGTCCCAAC 3780
Db 3721 GTACCCCGCAGGTTGAGGGCCAACTCTCCAGGGGACCATGAAGACGACCGTCCCAAC 3780
QY 3781 TTCTCCAGAGCTCCCGTTCCGTAGGGATGCTGAGCGGCAATCTCTCGAGAGT 3840
Db 3781 TTCTCCAGAGCTCCCGTTCCGTAGGGATGCTGAGCGGCAATCTCTCGAGAGT 3840
QY 3841 TCAGAGCTTCTGAGCGGCTCTCCAGCGGCTTGACATACATCTCCAAACGGGGGTTCA 3900
Db 3841 TCAGAGCTTCTGAGCGGCTCTCCAGCGGCTTGACATACATCTCCAAACGGGGGTTCA 3900
QY 3901 GCTGGTGTATAGGGGGGCCCCAAACAGAGGAAAGAAAGCCCTCATGTTCTCGGT 3960
Db 3901 GCTGGTGTATAGGGGGGCCCCAAACAGAGGAAAGAAAGCCCTCATGTTCTCGGT 3960
QY 3961 AAAGCAATCTAAGTCCCTTTTGGTATGAAGCCCTTCGAGGGGATTTTCGGCA 4020
Db 3961 AAAGCAATCTAAGTCCCTTTTGGTATGAAGCCCTTCGAGGGGATTTTCGGCA 4020
QY 4021 CCTCCATCTGGAGGGGGTCCGTTGGCCAAAGAAAGTCTCTGACCCCTATCTGACCCC 4080
Db 4021 CCTCCATCTGGAGGGGGTCCGTTGGCCAAAGAAAGTCTCTGACCCCTATCTGACCCC 4080
QY 4081 CTAGTGGCATCGGTGTGCTGGGTTTCTCTAAAGCTCTGTAAGCTTCAAGAG 4140
Db 4081 CTAGTGGCATCGGTGTGCTGGGTTTCTCTAAAGCTCTGTAAGCTTCAAGAG 4140
QY 4141 GTTTTTCTGTTCTTCAACCTCGGACCTCTTGTCTATCTGGAGCCCGAGGCTTACCTAG 4200

Db 4141 GTTTTTCGTTCTTCAACCTCGACCTCTGTCTCTGAGCCCGAGCGGTACCTAG 4200
Qy GTCTGGGGTGATCCGGGGCAACCGCTCGGTTTCGCTTTTATATGGGTCCAAATAAC 4260
Db GTCTGGGGTGATCCGGGGCAACCGCTCGGTTTCGCTTTTATATGGGTCCAAATAAC 4260
Qy CGTCAGCCAGCGGTGGCAATCCCGCTCTCTAAAGGCGGTATAGGCCCTGCTAGGAG 4320
Db CGTCAGCCAGCGGTGGCAATCCCGCTCTCTAAAGGCGGTATAGGCCCTGCTAGGAG 4320
Qy GGGGTAGTACTTTCTTACCCCTTAGGCTTGAGAGGCTTAGAGGCTCTCTAGGGCC 4380
Db GGGGTAGTACTTTCTTACCCCTTAGGCTTGAGAGGCTTAGAGGCTCTCTAGGGCC 4380
Qy TCCTGGGGTGAGGGGTAACTCATGSCCAGCCCGCGGCTCGGGACTCTGAGGAGG 4440
Db TCCTGGGGTGAGGGGTAACTCATGSCCAGCCCGCGGCTCGGGACTCTGAGGAGG 4440
Qy CCTCCATAGCTACTCGTGTGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 4500
Db CCTCCATAGCTACTCGTGTGAGGTTTGTGAAGGGGTTCACTAATGCATACGGCTAGC 4500
Qy CTGGGATCAGGGCCAAATGATGAGGTTTGGTATAAACCTCAGGTTGAGGCTA 4560
Db CTGGGATCAGGGCCAAATGATGAGGTTTGGTATAAACCTCAGGTTGAGGCTA 4560
Qy GTTTATGTGGTTTATGACCTTTACCTCGATCAGGGGCAACACACGAGTTTCTGTC 4620
Db GTTTATGTGGTTTATGACCTTTACCTCGATCAGGGGCAACACACGAGTTTCTGTC 4620
Qy ACAGAAAGAAACTTCCGATCTAAGAGGGGAAAGAGGTGAGAGGACGCGCTTCATG 4680
Db ACAGAAAGAAACTTCCGATCTAAGAGGGGAAAGAGGTGAGAGGACGCGCTTCATG 4680
Qy AAGTTTGGCTCTTAGAGGCGTGTCTAGAGGCGGTCTCGGGTTCAATCTTCCCTC 4740
Db AAGTTTGGCTCTTAGAGGCGGTGTAGAGGCGGTCTCGGGTTCAATCTTCCCTC 4740
Qy TCTCTCCAGGTTTCCGAGGTTTCGAGTCTTGTGTCAGGTTCTGATCAAGTTTTCAGCA 4800
Db TCTCTCCAGGTTTCCGAGGTTTCGAGTCTTGTGTCAGGTTCTGATCAAGTTTTCAGCA 4800
Qy AAGTCTATTCTCGGAATATAGGGGTATCTGTCTATCTTCCCTACGGGATATCTCTGCT 4860
Db AAGTCTATTCTCGGAATATAGGGGTATCTGTCTATCTTCCCTACGGGATATCTCTGCT 4860
Qy GTGTGAATGATCCCATCCCAATCATATCTCAATCTCTCTCTCTCTCTCTCTCTCT 4920
Db GTGTGAATGATCCCATCCCAATCATATCTCAATCTCTCTCTCTCTCTCTCTCTCT 4920
Qy ATCCCTAAATCT 4980
Db ATCCCTAAATCT 4980
Qy CGACCAAGAGAGCTTCTCGGGTCTAGTTTCGATATCTCGGACAGGTTTTCATCGTCT 5040
Db CGACCAAGAGAGCTTCTCGGGTCTAGTTTCGATATCTCGGACAGGTTTTCATCGTCT 5040
Qy AGACAGAGATTAGGCATGAAATAATGGCTTTGACAAATCTTTCTTAAATAATCTCCC 5100
Db AGACAGAGATTAGGCATGAAATAATGGCTTTGACAAATCTTTCTTAAATAATCTCCC 5100
Qy CGAGTTGGGAGTCCCTCGGGAGAGATTTTGGCAGTTTATGATGTTATGCTCTAT 5160
Db CGAGTTGGGAGTCCCTCGGGAGAGATTTTGGCAGTTTATGATGTTATGCTCTAT 5160
Qy CACGGCCCGAGGCTCCACGATAAGTTCTTTGGCCAAAGTACCGGGCCAGGTCGGGGT 5220
Db CACGGCCCGAGGCTCCACGATAAGTTCTTTGGCCAAAGTACCGGGCCAGGTCGGGGT 5220
Qy GCTCTTCAGGTGGTATGGTACTTTCAGGAGTTTCAAGTCTCTTTTAGAGGCTTCAG 5280

Db 5221 GCTCTTCAGCGTGGTGATGTTTCTTTCACGGAAGTTTCAAGTCTCTTTTAGAGGCTTCAG 5280
Qy GTCCGGGATAGTCTCAAGTACTCCCAAGCGTTTCTCGGCCCGTGGTTCGGGGAGAGGAC 5340
Db GTCCGGGATAGTCTCAAGTACTCCCAAGCGTTTCTCGGCCCGTGGTTCGGGGAGAGGAC 5340
Qy AAAGGGTTCGGGCAAAAGTTTCATCTTTGTACTTAGACCGGATTTAGCACCTGATAA 5400
Db AAAGGGTTCGGGCAAAAGTTTCATCTTTGTACTTAGACCGGATTTAGCACCTGATAA 5400
Qy CTTCAGGCGCTTAAAGGGCTCTCCTCGGAGACGCGTGGAGGAGCGTGGGGTG 5460
Db CTTCAGGCGCTTAAAGGGCTCTCCTCGGAGACGCGTGGAGGAGCGTGGGGTG 5460
Qy GAAAGACGAAACCCCGATTTTGGGAAGTCTCCCTCCAGTTTGTATGATGAACGTTGGGA 5520
Db GAAAGACGAAACCCCGATTTTGGGAAGTCTCCCTCCAGTTTGTATGATGAACGTTGGGA 5520
Qy GGAAGCGGCGGAGATGCTTTTCATCGGCTCGAACCTCGGACACATATAAAACCTTTCG 5580
Db GGAAGCGGCGGAGATGCTTTTCATCGGCTCGAACCTCGGACACATATAAAACCTTTCG 5580
Qy TGTTCGTCGGGCAAGAGTCTATGATAGAGTAACTTCGGGAGTACAAAGTGCCTCAA 5640
Db TGTTCGTCGGGCAAGAGTCTATGATAGAGTAACTTCGGGAGTACAAAGTGCCTCAA 5640
Qy GCCGCTTTTCCCAACGCTCCAAACCTCTAGGTCAGTGGTGGTTCCTGAAAACT 5700
Db GCCGCTTTTCCCAACGCTCCAAACCTCTAGGTCAGTGGTGGTTCCTGAAAACT 5700
Qy CTAGCTTTTCAGTGGTCAATCTCCTCACCCCTTAGCAGTACTCTCGAGGTAACCTTTG 5760
Db CTAGCTTTTCAGTGGTCAATCTCCTCACCCCTTAGCAGTACTCTCGAGGTAACCTTTG 5760
Qy ACACAGCGGCAAGTCTAGCGTCTCCAGTCCAGTGGTCTCGGACGCGTGAAGAGGGA 5820
Db ACACAGCGGCAAGTCTAGCGTCTCCAGTCCAGTGGTCTCGGACGCGTGAAGAGGGA 5820
Qy GGGGCTTGGTGTAGAGACCCAGAGACCC 5849
Db GGGGCTTGGTGTAGAGACCCAGAGACCC 5849

RESULT 2
AAK01848/c
ID AAK01848 standard; DNA; 1242 BP.
XX AAK01848;
XX AC
XX DT 13-APR-1999 (first entry)
XX Thermus sp. Tsp45IM genomic DNA.
XX Thermostable enzyme; Tsp45IM; restriction endonuclease; methylase;
XX cleavage; ss.
XX Thermus sp.
XX US586422-A.
XX 02-FEB-1999.
XX 29-OCT-1997; 97US-00960756.
XX 29-OCT-1997; 97US-00960756.
XX (NEW) NEW ENGLAND BIOLABS INC.
XX Xu S, Wayne J;
XX WPI; 1999-141942/12.
XX P-PSDB; AAW92372.

PT DNA coding for Tsp45I restriction endonuclease and methylase from *Thermus*
 PT species Y845 - useful for recombinant production of thermostable enzyme
 PT in *Escherichia coli*.
 XX
 PS Disclosure; Fig 1A-B; 22pp; English.

XX This sequence represents a novel restriction endonuclease methylase gene,
 CC Tsp45IM which has been isolated from *Thermus* sp. Y845. This DNA is useful
 CC for producing a thermostable endonuclease which can be used in a variety
 CC of assays, such as PCR, where high temperatures cannot be avoided. Tsp45I
 CC cleaves a double-stranded DNA molecule before the first G residue of the
 CC sequences 5'-GTGAC-3' and 5'-GTGAC-3'. The endonuclease creates sticky
 CC ends in the form of 5 nucleotide, single-stranded 5' overhangs. The
 CC production of recombinant Tsp45I in *E. coli* is simpler and produces 3
 CC multiply 105 units per gram of wet cell, about a ten-fold increase over
 CC that prepared from Y845
 XX

SQ Sequence 1242 BP; 291 A; 338 C; 336 G; 277 T; 0 U; 0 Other;
 Query Match 21.2%; Score 1242; DB 2; Length 1242;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1609	CTAGAGCGGACACAAATCTCAAACTTTGTGTGTAGCTGGGAAATCCTCTACACCCCT	1668
DB	1242	CTAGAGCGGACACAAATCTCAAACTTTGTGTGTAGCTGGGAAATCCTCTACACCCCT	1183
QY	1669	TCTAGTGAAGCTTTGACCGCTCCAGGAGGATCTATGCCGATGATGCCGCTTTAA	1728
DB	1182	TCTAGTGAAGCTTTGACCGCTCCAGGAGGATCTATGCCGATGATGCCGCTTTAA	1123
QY	1729	GAGGGTGAGCTATAAGCGTAGTACCGAGCCCTGCGAAGGATCGAGCACTAAATCCC	1788
DB	1122	GAGGGTGAGCTATAAGCGTAGTACCGAGCCCTGCGAAGGATCGAGCACTAAATCCC	1063
QY	1789	CTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTCTCGTGGGGTA	1848
DB	1062	CTCGTTACTCCCTGTTTGGACGATGAGCTTGAGCATGTCCAGATTTTCTCGTGGGGTA	1003
QY	1849	TCGGGGTACGAGGATCCTTGAATGCTGCAACCTCTCGGAGCTTTCCCTCTTTTCAG	1908
DB	1002	TCGGGGTACGAGGATCCTTGAATGCTGCAACCTCTCGGAGCTTTCCCTCTTTTCAG	943
QY	1909	CGCATCCGAGCGTAAATTTTTCGCGGCACCCCGTTCTTTGACGAGCAATAAGCCC	1968
DB	942	CGCATCCGAGCGTAAATTTTTCGCGGCACCCCGTTCTTTGACGAGCAATAAGCCC	883
QY	1969	TTGAGCGTGTAGCTGTCAAGCTTCTCGGGGATAGCCCAATGCCGTCCAGAGGGGG	2028
DB	882	TTGAGCGTGTAGCTGTCAAGCTTCTCGGGGATAGCCCAATGCCGTCCAGAGGGGG	823
QY	2029	AAGTATTCCTCGCAAGGCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATCGAG	2088
DB	822	AAGTATTCCTCGCAAGGCTTCCGGTAGGGCCATCCTTGGTTTCTCCAGGAGCATCGAG	763
QY	2089	GGGATGTGTGTGTACCGTTCCCGTTCTCGTCTACAAAGGGGAAAGCCCTAGCGATCTC	2148
DB	762	GGGATGTGTGTGTACCGTTCCCGTTCTCGTCTACAAAGGGGAAAGCCCTAGCGATCTC	703
QY	2149	CTCTTCCGAATAGGGGCTAGCGATTCGTTCCAAACGTAGTCCCGCTTTTGGAGTAGAC	2208
DB	702	CTCTTCCGAATAGGGGCTAGCGATTCGTTCCAAACGTAGTCCCGCTTTTGGAGTAGAC	643
QY	2209	GAGGATCATGTCTTTTCGATCCGAGCCTTACGGGAAAGTTTTCGGATTGTAAGC	2268
DB	642	GAGGATCATGTCTTTTCGATCCGAGCCTTACGGGAAAGTTTTCGGATTGTAAGC	583
QY	2269	GATCGGGCGATATGTTTAAAGAGTTTCGGCGCCCAAGACCTCATCAGGATGAGCTT	2328
DB	582	GATCGGGCGATATGTTTAAAGAGTTTCGGCGCCCAAGACCTCATCAGGATGAGCTT	523
QY	2329	CACCTCGAACCCGATTTTCTCGTTATGTGAACGAAGATCAGTCTGAGTCCGCGCATCAG	2388

DB	522	CACCTCGAACCCGATTTTCTCGTTATGTGAACGAGATCAGTCTGAGTCCGCGCATCAG	463
QY	2389	CTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACCTCCACAACTGAGGACCATCGAGGT	2448
DB	462	CTCCCTGAGAAGTATCAAGCGCTCCCTCAGGAACCTCCACAACTGAGGACCATCGAGGT	403
QY	2449	GTCACTGTAGCCCAACTGACCGCTTTTGGCTGGCTGACGGTAGCAACGCGATCTGTTTC	2508
DB	402	GTCACTGTAGCCCAACTGACCGCTTTTGGCTGGCTGACGGTAGCAACGCGATCTGTTTC	343
QY	2509	ATCGCGCCCAACGAGAAATCTGTCGCGGTTTCCATAAGCGGGTCAATATAGCAACTG	2568
DB	342	ATCGCGCCCAACGAGAAATCTGTCGCGGTTTCCATAAGCGGGTCAATATAGCAACTG	283
QY	2569	GACCTTCCCGCATACCCACGAGCTCCCGGAGCATCCACCGAGAACCTGACCGTTTC	2628
DB	282	GACCTTCCCGCATACCCACGAGCTCCCGGAGCATCCACCGAGAACCTGACCGTTTC	223
QY	2629	CCCCAAAAAGTAGTGGCAATAGGATCAATCTCAAAAAGGGGGGCAATTTCCCGCTAGGAA	2688
DB	222	CCCCAAAAAGTAGTGGCAATAGGATCAATCTCAAAAAGGGGGGCAATTTCCCGCTAGGAA	163
QY	2689	GAGGAGGGTTCTTTTCGCAAAACAAAGTTGTGGGTGGGCTGATCAAGATCTCTCTC	2748
DB	162	GAGGAGGGTTCTTTTCGCAAAACAAAGTTGTGGGTGGGCTGATCAAGATCTCTCTC	103
QY	2749	ATCGCGTTTTCGCGGGTAGACCAACCTAAAGGGCGAAGTTCCGAGGTTTCGAGGCTTT	2808
DB	102	ATCGCGTTTTCGCGGGTAGACCAACCTAAAGGGCGAAGTTCCGAGGTTTCGAGGCTTT	43
QY	2809	CAAGGGGGCTTTTCGCGGTCAAAACGAGGTAGTACGGTTCAT	2850
DB	42	CAAGGGGGCTTTTCGCGGTCAAAACGAGGTAGTACGGTTCAT	1

RESULT 3

AA04666/c
 ID AA04666 standard; DNA; 1026 BP.

XX AA04666;
 XX
 DT 04-JUL-2001 (first entry)

Thermus replication protein, RepT encoding DNA from pTsp45S plasmid.
 DE
 DE Replication protein; RepT; partition protein; ParA; pTsp45S plasmid;
 KW kanamycin-resistance gene; thermophilic transformation; Ori;
 KW replication origin; ds.

XX Thermus sp.

Key	Location/Qualifiers
CDS	1..1026
FT	/tag= a
FT	/product= "Replication protein, RepT"
FT	/codon= (Seq:GTG, aa:Met)

XX US6207377-B1.

XX 27-MAR-2001.

XX 14-AUG-1998; 98US-00134246.

XX 14-AUG-1998; 98US-00134246.

XX (NEWE) NEW ENGLAND BIOLABS INC.

XX Wayne J, Xu S;

XX WPI; 2001-298939/31.

XX P-PSDB; AAE01000.

XX Cloning *Thermus* species (Ts) plasmid genes comprises transforming

PT Escherichia coli with cloned recombinant plasmid containing Ts and E.coli
PT origins of replication, isolating cloned recombinant plasmid from E.coli
PT and transforming Ts cell.
XX
PS Example 1; Fig 1; 32pp; English.
XX
CC The present DNA sequence encodes Thermus sp. replication protein, Rept
CC which is obtained from pTsp45S plasmid. The replication protein, Rept is
CC needed for thermophilic plasmid replication. The invention relates to
CC Thermus sp. replication protein Rept, partition protein Para and their
CC corresponding DNA molecules which relates to recombinant DNA molecules
CC encoding plasmid DNA replication origins in Thermus, as well as to
CC shuttle vectors which contain the same. The invention also relates to
CC method useful for cloning Thermus sp. plasmid genes which comprises
CC inserting plasmid DNA comprising a Thermus sp. origin of replication
CC (Ori) into a recombinant plasmid comprising a thermostable kanamycin-
CC resistance gene and an Escherichia coli Ori, to produce a cloned
CC recombinant plasmid. This cloned recombinant plasmid is transformed with
CC an E. coli. host cell, and E. coli. host cell cultured for the expression
CC of cloned recombinant plasmid. The cloned recombinant plasmid isolated
CC from E. coli host cell is then transformed with Thermus sp. host cell and
CC Thermus sp. host cell is cultured. Thus Thermus sp. plasmid genes are
CC cloned. These plasmid DNAs are used for thermophilic transformation
XX
SQ Sequence 1026 BP; 221 A; 323 C; 285 G; 197 T; 0 U; 0 Other;
Query Match 17.5%; Score 1026; DB 5; Length 1026;
Best Local Similarity 100.0%; Pred. No. 5.1e-273;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
3132 TCAGGAGGCTCTTCCAGGCTCTTACTAGTACGCTCCCGCTCGTGCAGATGCCATC 3191
DB 1026 TCAGGAGGCTCTTCCAGGCTCTTACTAGTACGCTCCCGCTCGTGCAGATGCCATC 967
QY 3192 CGTGTATCTCGATACCGCATAGGACGCAATAGATATTCGGGGAGTTCACC 3251
DB 966 CGTGTATCTCGATACCGCATAGGACGCAATAGATATTCGGGGAGTTCACC 907
QY 3252 CTGGGCCACAGCCACAGCAAGCCTGCATAGAAACGCTTTGAACGACGCTCATCAGGAG 3311
DB 906 CTGGGCCACAGCCACAGCAAGCCTGCATAGAAACGCTTTGAACGACGCTCATCAGGAG 847
QY 3312 ATCGGCAATGTAGTACGATAGGCTGTATAGGCGCGGAGTTTGGACGCTCCACCTC 3371
DB 846 ATCGGCAATGTAGTACGATAGGCTGTATAGGCGCGGAGTTTGGACGCTCCACCTC 787
QY 3372 GGGGAGGACACAGATAGGCGCCAGGCTCAACGGCCACGCTTCTGGTGTGGGATCACCTC 3431
DB 786 GGGGAGGACACAGATAGGCGCCAGGCTCAACGGCCACGCTTCTGGTGTGGGATCACCTC 727
QY 3432 TTTCCCTGAGCCACAGGACACGACGCTCCAGGGTGGGGCGGATTCGGTGTCTGTGTA 3491
DB 726 TTTCCCTGAGCCACAGGACACGACGCTCCAGGGTGGGGCGGATTCGGTGTCTGTGTA 667
QY 3492 GGCCTTTGACCCAGTTCAAGAGACGACGCTTGGCCATGCTAGGGGAGGTTCTTCCA 3551
DB 666 GGCCTTTGACCCAGTTCAAGAGACGACGCTTGGCCATGCTAGGGGAGGTTCTTCCA 607
QY 3552 GGGGTAGATGTATGCTCCAGGGTGTAGCTGTCTTCCCTGCGCTCAGCCGACGGCCCA 3611
DB 606 GGGGTAGATGTATGCTCCAGGGTGTAGCTGTCTTCCCTGCGCTCAGCCGACGGCCCA 547
QY 3612 AAGGTTGGCGATGGCCCGGCTCCCGTTGACGTTTGGTGAAGAGCTGGTGGCCAC 3671
DB 546 AAGGTTGGCGATGGCCCGGCTCCCGTTGACGTTTGGTGAAGAGCTGGTGGCCAC 487
QY 3672 CAGGCCCTTTTCTCAAGGACCTTCTTCCAGCGCTGGAGCGGTCTCGGGGTGACCCCCAG 3731
DB 486 CAGGCCCTTTTCTCAAGGACCTTCTTCCAGCGCTGGAGCGGTCTCGGGGTGACCCCCAG 427
QY 3732 GTTAGGGGCAACATCTCCAGGGGACCATGAGACGACGCTCCCACTTCTCCACAG 3791
DB 426 GTTAGGGGCAACATCTCCAGGGGACCATGAGACGACGCTCCCACTTCTCCACAG 367

QY 3792 CTCGCCGTTCCCTAGGGATGCTGGAGCGGCAATCTCTCGAGGAGTTCCAGAGCTT 3851
DB 366 CTCGCCGTTCCCTAGGGATGCTGGAGCGGCAATCTCTCGAGGAGTTCCAGAGCTT 307
QY 3852 CTGAGGCCCTCTCTGGACGGCTTTGACATACATCCCAAGGGGGTTCAGCTGGTGGTAT 3911
DB 306 CTGAGGCCCTCTCTGGACGGCTTTGACATACATCCCAAGGGGGTTCAGCTGGTGGTAT 247
QY 3912 AGGGGGGGCCCCAACAGAGAGAAAGAAAGCTCATGGTCTTCTCGGGTAAGCAATCCT 3971
DB 246 AGGGGGGGCCCCAACAGAGAGAAAGAAAGCTCATGGTCTTCTCGGGTAAGCAATCCT 187
QY 3972 AAGTGCCTCTTTTGGTATGTAAGGCCCTTCGGAGGCGATTTTCGGCACCTCCATCTGG 4031
DB 186 AAGTGCCTCTTTTGGTATGTAAGGCCCTTCGGAGGCGATTTTCGGCACCTCCATCTGG 127
QY 4032 AGGGGGGTCGGTGGCCCAAGAAAGTCTCTGACCCCTTATCTGACCCCTTAGTGCCATC 4091
DB 126 AGGGGGGTCGGTGGCCCAAGAAAGTCTCTGACCCCTTATCTGACCCCTTAGTGCCATC 67
QY 4092 GGTGTTGTCGGTGGTTCCTCTAAAGCTCTGTAAGCTCTTCAAGAAAGTTCCTTCGTT 4151
DB 66 GGTGTTGTCGGTGGTTCCTCTAAAGCTCTGTAAGCTCTTCAAGAAAGTTCCTTCGTT 7
QY 4152 CTTTAC 4157
DB 6 CTTTAC 1
RESULT 4
AA01849
ID AA01849 standard; DNA; 999 BP.
XX
AC AA01849;
XX
DT 13-APR-1999 (first entry)
XX
DE Thermus sp. Tsp45I genomic DNA.
XX
KW Thermostable enzyme; Tsp45I; restriction endonuclease; cleavage; ss.
XX
OS Thermus sp.
XX
FN US5866422-A.
XX
PD 02-FEB-1999.
XX
PF 29-OCT-1997; 97US-00960756.
XX
PR 29-OCT-1997; 97US-00960756.
XX
PA (NEWE) NEW ENGLAND BIOLABS INC.
XX
PI Xu S, Wayne J;
XX
DR WPI; 1999-141942/12.
DR P-PSDB; AAW92373.
XX
PT DNA coding for Tsp45I restriction endonuclease and methylase from Thermus
PT species YS45 - useful for recombinant production of thermostable enzyme
PT in Escherichia coli.
PS Disclosure; Fig 2A-B; 22pp; English.
XX
CC This sequence represents a novel restriction endonuclease gene, Tsp45I
CC which has been isolated from Thermus sp. YS45. This DNA is useful for
CC producing a thermostable endonuclease which can be used in a variety of
CC assays, such as PCR, where high temperatures cannot be avoided. Tsp45I
CC cleaves a double-stranded DNA molecule before the first G residue of the
CC sequences 5'-GTGAC-3' and 5'-GTGAC-3'. The endonuclease creates sticky
CC ends in the form of 5 nucleotide, single-stranded 5' overhangs. The
CC production of recombinant Tsp45I in E. coli is simpler and produces 3

CC multiply 105 units per gram of wet cell, about a ten-fold increase over
CC that prepared from YS45
XX SQ Sequence 999 BP; 310 A; 219 C; 299 G; 171 T; 0 U; 0 Other;

Query Match 17.1%; Score 999; DB 2; Length 999;
Best Local Similarity 100.0%; Pred. No. 1.5e-265;
Matches 999; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 614 ATGCAACAGATGGCGGAGTGAAGCTGTGGACACAGAGAAGCGTTGAGCTTCTTGGAGAAG 673
DB 1 ATGCAACAGATGGCGGAGTGAAGCTGTGGACACAGAGAAGCGTTGAGCTTCTTGGAGAAG 60

QY 674 GGGTATTGGATAAATACTACAGCTCTATAAAGGGGAAAGTGGCTTCTGAGTCACTA 733
DB 61 GGGTATTGGATAAATACTACAGCTCTATAAAGGGGAAAGTGGCTTCTGAGTCACTA 120

QY 734 CCAGAGAGGTAGAGGAAATACTTCGCGAGCGCTTACAGGCATACAGAGGGGAGCGAGAT 793
DB 121 CCAGAGAGGTAGAGGAAATACTTCGCGAGCGCTTACAGGCATACAGAGGGGAGCGAGAT 180

QY 794 AGTCGCGAGGAGAGAAACGAAACTCGTGAAGCGCTGTAAATGCCAGAAAAAAGGTCGAG 853
DB 181 AGTCGCGAGGAGAGAAACGAAACTCGTGAAGCGCTGTAAATGCCAGAAAAAAGGTCGAG 240

QY 854 CGGTCCCGCTTCAATACCCCTACTGCTTGTCTTACTACTGCTTGTGGAAGAAGCA 913
DB 241 CGGTCCCGCTTCAATACCCCTACTGCTTGTCTTACTACTGCTTGTGGAAGAAGCA 300

QY 914 GAAAGAGGAAACAGGCGCTTGAAGGAGCATTTGAGAGGTTGCTCAAAGCAGCCAGAA 973
DB 301 GAAAGAGGAAACAGGCGCTTGAAGGAGCATTTGAGAGGTTGCTCAAAGCAGCCAGAA 360

QY 974 ACCATCCGCTGCTTGGCCAAAGGAGCGCAAGAGAGCGGTAGAAGCCTTGTATCCAAAGG 1033
DB 361 ACCATCCGCTGCTTGGCCAAAGGAGCGCAAGAGAGCGGTAGAAGCCTTGTATCCAAAGG 420

QY 1034 CTCAGAGGAGCTCCGGAATAAATCGGCAGATAGCGCGCATTTCAAAGGTGTACAAA 1093
DB 421 CTCAGAGGAGCTCCGGAATAAATCGGCAGATAGCGCGCATTTCAAAGGTGTACAAA 480

QY 1094 GAAGAGCTAAAGGGGAAATAAGAGAGGCTTCCAGGCGCTTACCAACCAAGAGATTGTG 1153
DB 481 GAAGAGCTAAAGGGGAAATAAGAGAGGCTTCCAGGCGCTTACCAACCAAGAGATTGTG 540

QY 1154 GTAGTATCCCTGAAAGAGTAACCGGAGCAAGCAGCCCTTATTGCGGAGAGAGCG 1213
DB 541 GTAGTATCCCTGAAAGAGTAACCGGAGCAAGCAGCCCTTATTGCGGAGAGAGAGCG 600

QY 1214 GGCAATCATATATACCGGATCGGATGAAGCTTTGAAAGATGCGCGCAAGGAAACCTG 1273
DB 601 GGCAATCATATATACCGGATCGGATGAAGCTTTGAAAGATGCGCGCAAGGAAACCTG 660

QY 1274 GGCCTTGGCGAGGAGAGACTAGGCAACCAAGGCGTAGATTTCTAGTGGTCAATCGG 1333
DB 661 GGCCTTGGCGAGGAGAGACTAGGCAACCAAGGCGTAGATTTCTAGTGGTCAATCGG 720

QY 1334 CGTAGCCCTGAGAGACATGCACTACAGAGAGTGAAGTTTCAATCCGACTTTGGC 1393
DB 721 CGTAGCCCTGAGAGACATGCACTACAGAGAGTGAAGTTTCAATCCGACTTTGGC 780

QY 1394 GGAAACCAAGACAAACAGAAACTAGTAGCAAGGCTTCCATAAGGTTGGACCTTTGAGAAG 1453
DB 781 GGAAACCAAGACAAACAGAAACTAGTAGCAAGGCTTCCATAAGGTTGGACCTTTGAGAAG 840

QY 1454 AGGCATATAGATAGTGGTGGAGCGGATGCTGTGGTGAACAAGTTTCGTGGGTGG 1513
DB 841 AGGCATATAGATAGTGGTGGAGCGGATGCTGTGGTGAACAAGTTTCGTGGGTGG 900

QY 1514 GCCGAGCTGGGGAAGAAACCATCGTTACATCCGCTACTCCTCTCCAGACCTGATAGCG 1573
DB 901 GCCGAGCTGGGGAAGAAACCATCGTTACATCCGCTACTCCTCTCCAGACCTGATAGCG 960

QY 1574 GAGCTCTACCAAAAGGGTGAGAGAGCCCTGGGCTCTTAG 1612
DB 961 GAGCTCTACCAAAAGGGTGAGAGAGCCCTGGGCTCTTAG 999

RESULT 5
AAT67772/c
ID AAT67772 standard; DNA; 1032 BP.
XX AAT67772;
AC
XX 29-JUL-1997 (first entry)
DT
XX H. pylori cytoplasmic protein ORF 24816915.aa.
DE
XX Vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; cytoplasmic; ds.
XX
OS Helicobacter pylori.
XX

Key Location/Qualifiers
CDS 1..1032
FT /*tag= a
FT
XX
PN WO9640893-Al.
XX
PD 19-DEC-1996.
XX
XX 06-JUN-1996; 96WO-US009122.
XX
XX 07-JUN-1995; 95US-00487032.
XX 01-APR-1996; 96US-00630405.
XX
XX (ASTR) ASTRA AB.
XX
XX Smith D, Berglindh OT, Mellgaerd BL;
XX WPI; 1997-052306/05.
XX P-PSDB; AAW20333.
XX
XX Helicobacter pylori nucleic acid sequences and related polypeptide(s) -
XX useful for vaccines to treat or prevent H. pylori infection, and to
XX detect Helicobacter.
XX
XX Claim 9; Page; 1481pp; English.
XX
XX The present sequence encodes a Helicobacter pylori cytoplasmic protein.
XX The protein may be used in a vaccine to prevent or treat H. pylori
XX infection or to identify H. pylori polypeptide binding compounds, useful
XX as potential H. pylori life cycle activators or inhibitors. The genomic
XX sequence of H. pylori (ATCC 55679) was determined from overlapping
XX contigs generated by mechanically shearing the bacterial DNA. The
XX sequences were analysed for ORF of at least 180 nucleotides, and the
XX predicted coding regions defined by computer evaluation. To identify
XX likely H. pylori antigens for vaccine development, the amino acid
XX sequences predicted from various ORF were analysed for significant
XX homology to other known or exported membrane proteins. Having identified
XX and determined the sequences of interest, particular regions can be
XX isolated from H. pylori by PCR amplification for recombinant polypeptide
XX production, e.g. in E. coli hosts. Note: This DNA sequence is not
XX reproduced in the specification and has been derived from the related
XX specification, WO9719098
XX
XX Sequence 1032 BP; 386 A; 166 C; 170 G; 310 T; 0 U; 0 Other;

Query Match 3.2%; Score 185.6; DB 2; Length 1032;
Best Local Similarity 52.0%; Pred. No. 3e-40;
Matches 4/5; Conservative 0; Mismatches 424; Indels 15; Gaps 2;

QY 1664 ACCCTTCTAGTGAAGCTTTGACCGCTCCCGAGGAGCATCTATGCGATGGATCCCGC 1723
DB 971 AGCTTGTTTTTGAAGCTTTGGATAGCAAAATCGAATTATCAATGCTATATAATTTTCTG 912

KW	bacterial life cycle; vaccine; immunisation; detection; antisense; inhibition; cytoplasmic; ds.
OS	Helicobacter pylori.
FH	Key
FT	Location/Qualifiers
FT	CDS
XX	1..1032
XX	/*tag= a
XX	WO9719098-A1.
XX	29-MAY-1997.
XX	15-NOV-1996;
XX	96WO-US018542.
XX	17-NOV-1995;
XX	95US-00561469.
XX	(ASTR) ASTRA AB.
PA	Smith DH;
XX	WPI; 1997-298052/27.
PI	P-PSDB; AAW24634.
DR	
DR	
XX	Helicobacter pylori nucleic acid sequences and related proteins - used for diagnostics and therapeutics.
PT	
PS	Claim 1; Page 100; 235pp; English.
XX	
CC	The present sequence encodes a Helicobacter pylori cytoplasmic protein. H. pylori has been strongly linked to chronic gastritis and duodenal ulcer disease. The nucleic acid sequences of the invention are used to evaluate compounds, especially activators or inhibitors of bacterial life cycle, for the ability to bind an H. pylori nucleic acid sequence. The nucleic acid sequences, and corresponding proteins, are also useful for generating vaccines for immunising subjects against H. pylori or for use in detecting the presence of Helicobacter species in a sample. Antisense nucleic acid sequences of these sequences are used to inhibit expression of a gene from Helicobacter species. H. pylori whole genomic DNA was isolated and nebulised to a median size of 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique BstXI-linker adapters in 100-1000 fold molar excess. These linkers are complementary to the BstXI-cut pmx vectors, while the overhang is not self-complementary. Therefore the linkers will not concatamerise nor will the cut vector re-ligate itself easily. The linker-adaptor inserts were ligated to each of the 20 pmx vectors to construct a series of shotgun subclone libraries. The purified DNA samples were then sequenced. Note: The ORF/protein reference number for this sequence was obtained from the related specification, WO9640893
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	
CC	

Db 735 --AACTGTTCTAAGCATAAATTTTTCCTAGGGTTATTTGTTAGATACTCAATC 678
Qy 1964 AGCCCTTGAGCGTCTAGCTCGTCAAGCTTCTCCGGGGATAGGCCCAATCCGTCAGGA 2023
Db 677 AAACCTTTCTTTATCCCAACGCTCAAGTGTGGCAATATCAGTGGCCCAATCCGCCCTTTT 618
Qy 2024 GGGGGAAGTATTCCTCGCAAGGCTTCGGGTAGGGCCATCCTTGTTTCTCCAGGAGCA 2083
Db 617 GGAGGTAGCATACCTTTAAATGCTTTTAGACATTCGCCACTTTCCACTTCTCTCGGAGCA 558
Qy 2084 TGCAGGGATGTTGGTGTACCGTTCCCGGTTCTCGTCTCAAAAGGGGAAAGCCTTAGCG 2143
Db 557 TGTATTGGAACGTTAGTGTACCGCTTTTATCTTTGTCAATTTTAGGGAATCGTTTCT 498
Qy 2144 ATCTCCTCTTCGCAATAGGGCTAGCGGATTCGTTCCAAAGTAGTCCCGGTTTTCGAG 2203
Db 497 AAATCTTGTGCGGTATAGGATCTTAGTTCGTTAAATAGGATTTTTCCTTTAGAG 438
Qy 2204 TAGACGAGGATCATGCTCTTTTGGATCCGAGCCCTTACGGAAAGTTTTCGGGATTT 2263
Db 437 TAAATAAAATCATATCTTTTATGTTACCATAGCTATCCTTTTAAATTTTTCAGGATTTG 378
Qy 2264 GAACGATCGCGGATATGTTTAAACGAGTTTCGCGGCAAGACCTCATCAAGATG 2323
Db 377 CACTTTATGCGGTGATTCATTTCTAAATTTTGTATGCCAATATTTTCATCTAACATT 318
Qy 2324 AGCTTCACTCGAACCCGCTATTTCTCGTCTATGTGAACGAAGATCAGTCTGATCGGCC 2383
Db 317 ACCTTGACATAATGCCCTATCTTGAATCTGTATGCATAGATAGACCTTGTTCGAA 258
Qy 2384 ATCAGTCCCTGAGAAGTATCAAGCGTCCCTCAGGAATCCCAAACTGAGGACCATCG 2443
Db 257 AGCAATCTTTAAGCAATACCAAGCGTTGTTTAAAAATTTCCATAAAATCCATACCCACT 198
Qy 2444 AGGTGTCTATGTCAGCCCACTGACCGTTTTCGGCTGGCTAGCGTAGCAACGCCATCT 2503
Db 197 ACTTTACATATAGCAATATCCCATCTTAGAATGCTAATGTGTTGCTCTACCA 138
Qy 2504 GTTTTCATCGCGCAACGAGAAATGCTGCGCGTTCCATAGCGCGGTCAATATAGAC 2563
Db 137 TTT-----GTGATAGTAAATGATTTGTTCTAGCAATAGGAGGTCAATATATATC 87
Qy 2564 AACTGGACCTTCCC 2577
Db 86 AAATCTATTTTCC 73

RESULT 7

ID AAD04667/c

XX AAD04667 standard; DNA; 180 BP.

AC AAD04667;

XX 04-JUL-2001 (first entry)

DT Thermus promoter sequence upstream of replication protein rept gene.

XX Replication protein; RepT; partition protein; ParA; pTsp45S plasmid;

KW kanamycin-resistance gene; thermophilic transformation; Ori;

KW replication origin; promoter; ds.

XX Thermus sp.

OS

XX Key Location/Qualifiers

FH promoter 1..150

FT -35_signal /*tag= a

FT -10_signal /*tag= b

FT -35_signal /*tag= c

FT -35_signal /*tag= d

FT

-10_signal 109..114
FT /*tag= e
FT 139..143
FT /*tag= f
FT /standard_name= "Shine-Dalgarno Sequence"

PN US6207377-B1.

XX 27-MAR-2001.

XX 14-AUG-1998; 98US-00134246.

XX 14-AUG-1998; 98US-00134246.

PR (NEW) NEW ENGLAND BIOLABS INC.

XX Wayne J, Xu S;

XX WPI; 2001-298939/31.

XX Cloning Thermus species (Ts) plasmid genes comprises transforming

PT Escherichia coli with cloned recombinant plasmid containing Ts and E.coli
PT origins of replication, isolating cloned recombinant plasmid from E.coli
PT and transforming Ts cell.

XX Example 1; Fig 2; 32pp; English.

CC The present DNA sequence is Thermus sp. promoter sequence upstream of
CC replication protein, rept gene. The invention relates to Thermus sp.
CC replication protein RePT, partition protein ParA and their corresponding
CC DNA molecules which relates to recombinant DNA molecules encoding plasmid
CC DNA replication origins in Thermus, as well as to shuttle vectors which
CC contain the same. The invention also relates to method useful for cloning
CC Thermus sp. plasmid genes which comprises inserting plasmid DNA
CC comprising a Thermus sp. origin of replication (Ori) into a recombinant
CC plasmid comprising a thermostable kanamycin-resistance gene and an
CC Escherichia coli Ori, to produce a cloned recombinant plasmid. This
CC cloned recombinant plasmid is transformed with an E. coli. host cell, and
CC E. coli. host cell cultured for the expression of cloned recombinant
CC plasmid. The cloned recombinant plasmid isolated from E. coli host cell
CC is then transformed with Thermus sp. host cell and Thermus sp. host cell
CC is cultured. Thus Thermus sp. plasmid genes are cloned. These plasmid
CC DNAs are used for thermophilic transformation

XX Sequence 180 BP; 48 A; 44 C; 56 G; 32 T; 0 U; 0 Other;

Query Match 3.1%; Score 180; DB 5; Length 180;

Best Local Similarity 100.0%; Pred. No. 4e-39;

Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4128 CTCTCAAAAGAGTTTTCGTTCTTACCCCTCGGACCTCCTTGTCTCATCTGGAGCCGA 4187

Db 180 CTCTCAAAAGAGTTTTCGTTCTTACCCCTCGGACCTCCTTGTCTCATCTGGAGCCGA 121

Qy 4188 GCGTTTACCTAGTCTTGGGGTGATCCGGGCAACCCGCTCGGTTTCGCTTTTATG 4247

Db 120 GCGTTTACCTAGTCTTGGGGTGATCCGGGCAACCCGCTCGGTTTCGCTTTTATG 61

Qy 4248 GTTCCAAATACCGTCAGCCAGCGCTGGCAATCCCCCTCCCTAAAGGCGTTATAG 4307

Db 60 GGTCCAAATACCGTCAGCCAGCGCTGGCAATCCCCCTCCCTAAAGGCGTTATAG 1

RESULT 8

ADA71938/c

ID ADA71938 standard; DNA; 2000 BP.

XX ADA71938;

XX ADA71938;

DT 20-NOV-2003 (first entry)

XX Rice gene, SEQ ID 5263.

XX

KW Plant; bacterial infection; fungal infection; viral infection; rice;
 KW gene; ds.
 XX
 OS Oryza sativa.
 XX
 PN W02003000898-A1.
 XX
 XX 03-JAN-2003.
 XX
 XX 22-JUN-2001; 2001WO-IB001105.
 XX
 XX 22-JUN-2001; 2001WO-IB001105.
 XX
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
 PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
 XX
 XX WPI; 2003-175290/17.
 DR
 XX
 XX Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 PT gene expression.
 XX
 XX Claim 27; SEQ ID NO 5263; 899pp; English.
 XX
 XX The present invention relates to a method (M1) for identifying genes
 CC involved in plant resistance or response to pathogenic infection. M1
 CC comprises identifying a gene whose expression is significantly altered in
 CC the incompatible interaction of plant gene expression relative to
 CC expression of the gene in an uninfected plant, in a mutant plant that
 CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance to resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.
 XX
 XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
 SQ
 Query Match 1.0%; Score 59; DB 7; Length 2000;
 Best Local Similarity 8.4%; Pred. No. 5.1e-05;
 Matches 61; Conservative 339; Mismatches 324; Indels 1; Gaps 1;
 QY 2924 GCCAAGGCGGCTACCTTTGAGCCGATCTTCCCTGAGCTATAGACCTTCGATCG 2983
 DB 746 KSTRWGMGRMKSARWYCSRNKCAKTYASARWTKAKRSYRIRRWYWRKRGTYR 687
 QY 2984 TCTCAGGGTGACCCGAGGATGTCTGCAAGCTCTCGGGGTGAGGTACACGGGCTCA 3043
 DB 686 YRYSRCRTRARMSRRKRWAGSMKSCWYWRGARSMMYSKCSAKCKXRYMTSSY 627
 QY 3044 TCTCATGACACACTTACCCACAGAGACACACATGCAACTATGGGCAAGTAGAC 3103
 DB 626 MSTGMYTSSYSKMSWTSMYKMTCTMTYKMGSTRSRKMGWSGMSRMYRWWKX 567
 QY 3104 AACGACACAAAGCTTGGGCCACTCTCTCAGAGGCGCTCTTGGGGTCTTCACTAGT 3163
 DB 566 MRKRYMYRMYKWTWRWCYRWGTYMTTSSRMWYTGKRYTSKRYMYKRYKY 507
 QY 3164 AGCTCCGGTGCTGATGCGCATTCGGTGTATCTCGGATTAACCCGATGACGG 3223
 DB 506 CWYTYGMYTKSMYRMYGCKACKCCYAMCKWAKAYSGMMYMYRKYSKMRNSTKYM 447
 QY 3224 CAATAGATATTCGGGGGAGTTCACCTCGCCACACAGCCACAGCAAGCTGATAGA 3283
 DB 446 SMWYKCRNKGAKCYCKWYTCYGYMKNYIMSGYSKYRKYMYRMYKMYM 387
 QY 3284 AACGTTTGAACGACGGTCACTAGAGATCGGCAATGATGATCAATGAGGGTGATA 3343
 DB 386 MYYSASSMWTWYTYAKYWKYRRGTMSWYKSYKKYCTWYCMKRCYRWRKM 327
 QY 3344 GGGCCGGGAGTTTGAACGCTCCACCTCGGGGAGGACAGGATGAGGCCAGGTCAACGG 3403

DB 326 RKTKYKRCYCWRYATC-YWCCCYRKRGMYSRRSMRTAGKWKMRWSRWCFSYSWYKM 268
 QY 3404 CCAGGTCCTTGGTGTGGCATCACCTTTTCCCTGAGCCAGAGGACGACGACCTCCA 3463
 DB 267 YKQWKXSYMYSGWARSSTGTSAAKTYKYSISRRAKWRACRMYISACRRYRTSY 208
 QY 3464 GGGTGGGCGGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3523
 DB 207 YGCSYCGSKWYKSKYSCSRMTCSWCCCTCYGACWCSGCMWMTMGCGCVTRG 148
 QY 3524 TGGCCATGCTAGGGCGAGGTTCTCCAGGGGTAGATGCTGCTCAGGGTGCAGCTGG 3583
 DB 147 WKRSKYSKMCCKKYCSCCTKYCYRYCKWYKYSYKCYCYCYWYMSYMYRMYKCM 88
 QY 3584 CTTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3643
 DB 87 CSRSKSSWMSCAVCTSTSTSRWMSMYAAKMGYCGSGMYRMSKSKMYSKYSCKYTG 28
 QY 3644 CGGT 3648
 DB 27 KKCTK 23
 RESULT 9
 ADA71938
 ID ADA71938 standard; DNA; 2000 BP.
 XX
 AC ADA71938;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Rice gene, SEQ ID 5263.
 XX
 KW Plant; bacterial infection; fungal infection; viral infection; rice;
 KW gene; ds.
 XX
 OS Oryza sativa.
 XX
 XX W02003000898-A1.
 XX
 XX 03-JAN-2003.
 XX
 XX 22-JUN-2001; 2001WO-IB001105.
 XX
 XX 22-JUN-2001; 2001WO-IB001105.
 XX
 XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
 PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
 XX
 XX WPI; 2003-175290/17.
 DR
 XX
 XX Identifying at least one gene involved in plant resistance or response to
 PT pathogenic infection for conferring resistance or tolerance to a plant to
 PT bacterial, fungal or viral infection by determining or detecting plant
 PT gene expression.
 XX
 XX Claim 27; SEQ ID NO 5263; 899pp; English.
 XX
 XX The present invention relates to a method (M1) for identifying genes
 CC involved in plant resistance or response to pathogenic infection. M1
 CC comprises identifying a gene whose expression is significantly altered in
 CC the incompatible interaction of plant gene expression relative to
 CC expression of the gene in an uninfected plant, in a mutant plant that
 CC does not express a gene associated with response to pathogenic infection,
 CC or in a corresponding incompatible or compatible interaction. (M1) is
 CC useful for conferring resistance to resistance or tolerance to a plant to
 CC bacterial, fungal or viral infection. The present sequence was used to
 CC illustrate the invention.
 XX
 XX Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
 SQ

```
Query Match      1.0%; Score 57.6; DB 7; Length 2000;
Best Local Similarity 10.6%; Pred. No. 0.00012;
Matches 97; Conservative 394; Mismatches 418; Indels 5; Gaps 2;

QY 621 AGATGGCGGAGTGGAGCGTGGACACAGAGCGGTGAGCTTCTGGAGAGGGGTATT 680
D 16 RRRRYTMAGMMSCARMSRSMKSRMSKRYKSCSGCKMTRKRSKYWSASASGR 75
QY 681 TGGATAAAGTACTAGCTAGCTTATAAGGGGAAAGTGGCTTTGAGGTGATACAGAGG 740
D 76 TGSXSSGSSGCKGKRYKESKWRGRGRMRMRMRWGRYRRCARSGRVAGSGRM 135
QY 741 AGGTAGAGGAAAACTTCGAGCGCTTACAGGCATACAGGCGGAGGAGGAGATAGTCGG 800
D 136 MGGSRMSYMWYARGCGSKKSKSGSGWGTCTCRGARGSGWSGAKYKSGMSKRM 195
QY 801 AGGCAAGAAAGAACTCTGGAAGCGCTGTAATGCGAGAAAAAGGTGAGCGGTCCC 860
D 196 WYSSCGRSGCGRSAYSPRYTSKYTYKQYYSASRCWRAVMTTYSWACSSYTWCR 255
QY 861 CCTCAATACCCCTACCTGCTTGGTCTACTACCTGTTTGGAAAAAGCAGAAAAAG 920
D 256 SKRSMWMMWRKRWRSYGMYSNSYMMWMTCTAYKSYSRWCYMYRGGWGATRYG 315
QY 921 CGAACAAGCGCTTGAGGAGCATTTGAGGAGGTGCTCAAGCACCAGCAACCAATCC 980
D 316 RGYSRMAMMYKMYWRYGKMRGHWAGWWRSMC---RWSKACYWWRWWRTRR 372
QY 981 CGGTCTCTGGCAAGAGCGCAAGAGCGGTAGAACCTTCAATCAAGGCTCAAGG 1040
D 373 RRAKKSRTSRKRWKRWKRYKRYRMYRSGYRMRKCRARWKRCSRGRWAWKRCRGM 432
QY 1041 AGCTCCCGAATAAATCGGAGATAGCGCGATGTTCAAAAGGTGTACAAAGAGC 1100
D 433 TCRMSYGMWMSKRWKRWASKYKMSRMYRWRKRCSTRTWGTGRTGMMTGRCRYK 492
QY 1101 TAAAGGGGAAATAGAGAGAGGTTCAGGCGCTTACCAACCAAGATTGTGTATAT 1160
D 493 RSGMKRCRRRRRWGRWYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 552
QY 1161 CCCCTGAAAAAGTAAACCGAGGAGCAGCACCCCTTATTCGGAGAGAGAGCGGATCA 1220
D 553 GMMWKRYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 612
QY 1221 TCATATACAGGATCGGATAGCTTTGAAGATGCGCGCAAGAAACCTGGGCGCTTG 1280
D 613 SMRSRCKRCKASKESSAKRYAMWGMGTSGSRMSRNSYTCYWRKWSMKSTCTMYMS 672
QY 1281 GCGAGAGAGCAACTAGGCAAGCGGTAGATTCTTCTAGTGTATATCGGCGGTAGCC 1340
D 673 KYTYAKYGSYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 732
QY 1341 CTGAGAGACATGGCACCTAACAGGAGAGTGAAGTTTCAATCCGACTTTTGGCGGAAAC 1400
D 733 WYKCKSKWYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMY 791
QY 1401 AAGCAACACAGAACTAGTAGCAAGGCTTCCATAAGGTGAGCCTTGAGAGAGGACCA 1460
D 792 -WRYTYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 850
QY 1461 TAGGAATAGTGGTGGACGAGATCCCTGTTGGTGGAGCAAGTTTCGTGGGTGGCCCGAC 1520
D 851 WKKWATCATKRWMTKRGKAWMTMAKRWKRYRMYRMYRMYRMYRMYRMYRMYRMY 910
QY 1521 TGGGAAAGAAACG 1534
D 911 RWGWAYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYRMYR 924

RESULT 10
ADA70342/c
ID ADA70342 standard; DNA; 2030 BP.
```

```
XX AC ADA70342;
XX DT 20-NOV-2003 (first entry)
XX DE Rice gene, SEQ ID 3665.
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;
XX OS Gene; ds.
XX OS Oryza sativa.
XX PN WO2003000898-A1.
XX PD 03-JAN-2003.
XX PF 22-JUN-2001; 2001WO-IB001105.
XX PR 22-JUN-2001; 2001WO-IB001105.
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,
PI Katagiri F, Qian S, Tao Y, Whittham S, Xie Z, Zhu T, Zou G;
XX WI MPI; 2003-175290/17.
XX PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX PS Claim 6; SEQ ID NO 3665; 899pp; English.
XX CC The present invention relates to a method (M1) for identifying genes
XX CC involved in plant resistance or response to pathogenic infection. M1
XX CC comprises identifying a gene whose expression is significantly altered in
XX CC the incompatible interaction of plant gene expression relative to
XX CC expression of the gene in an uninfected plant, in a mutant plant that
XX CC does not express a gene associated with response to pathogenic infection,
XX CC or in a corresponding incompatible or compatible interaction. (M1) is
XX CC useful for conferring resistance or tolerance to a plant to
XX CC bacterial, fungal or viral infection. The present sequence was used to
XX CC illustrate the invention.
XX SQ Sequence 2030 BP; 388 A; 659 C; 727 G; 246 T; 0 U; 10 Other;

Query Match      1.0%; Score 56.4; DB 7; Length 2030;
Best Local Similarity 46.7%; Pred. No. 0.00027;
Matches 214; Conservative 0; Mismatches 241; Indels 3; Gaps 1;

QY 3348 CGGAGTTTGGAAACGCTCCACCTCGGGAGGAGCAGGATGAGGCCAGTCAACGCCAC 3407
D 942 CGCGCGCTTGGCGCGCTCGCACTCGCGCGGAGCTTCCCGCAGCGCGCGCGCGCG 883
QY 3408 GGTCTGTGTGGGATCACCTTTTCCCTGAGCCAGAGACAGACAGTCCAGGTT 3467
D 882 GATGTCGCGCGCGCTGCTTCCGCGGATGACCTTGACGAAGTGGTCCATGAGGGTTGTC 823
QY 3468 GGGG---CGAATTCCTGGTCTCTGGTAGGCTTGACCCAGTTGAAGAGAGACACGCCGTT 3524
D 822 GAAGTCTCTCGCGCGCGAGGTGGGTGCGCGTTGGTGGCAAGACCTCGAACACGCCGTT 763
QY 3525 GGCATGTCTAGGGCGAGGTTCTCTCCAGGGTAGATGTAGTCTGTCAGGGTGAAGCTGGC 3584
D 762 GTCGATGGCGAGGATCTGACGTCGAAACGTCGCCCGCGCGAGGTGAAAGCAGGAGCGTT 703
QY 3585 TTTTCCCTGGCTCAGCGGACGCGCCCAAGGGGTGCGGATGGCCCGCGCTCCCGTTGAC 3644
D 702 CTTCTCGGCGCCCTTCTTGTGATGCGGTAGGCGATGGCGGCGGTGCGGCTCGTTGAT 643
QY 3645 GGTTCGTGAAGACGTCGGTGGCCACAGGCCCTTTTCTCAAGACCTTCTTCAGGC 3704
```


Db 642 GATGGGTGACGGTGAGCCGGCGGATGACGGCGGCTCTTGGTGGCTCGGCTGGCG 583
 QY 3705 GTGACGGTCTGCCGGTGTACCCCGAGGTGAGGGCAACATCTCCAGGGGACCATGAA 3764
 Db 582 GTGCTTGAAGTAGGCCGGGACGGTGACGACGGCGCGGTGACCTTCTCGCGAGGTAGGC 523
 QY 3765 GAGACGGTCCCACTTCTCCAGAGCTCCCGTTGC 3802
 Db 522 CTCGGCGTCTCTTCTATCCGGTGGACCATGGCGC 485

RESULT 11

ADC08176/c
 ID ADC08176 standard; DNA; 2031 BP.

XX AC ADC08176;

XX DT 18-DEC-2003 (first entry)

XX DE Rice DNA sequence Seq ID443 related to grain filling.

XX KW plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
 KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
 KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
 KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
 KW gene; ds; plant.

XX OS Oryza sativa.

XX PN WO200300905-A2.

XX PD 03-JAN-2003.

XX PF 21-JUN-2002; 2002WO-1B002450.

XX PR 22-JUN-2001; 2001US-0300112P.

XX PR 26-SEP-2001; 2001US-0325277P.

XX PR 20-DEC-2001; 2001US-0342327P.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;

XX PI Glazebrook J, Katagiri F, Kreps J, Provart N, Rieke D;

XX PP WPI; 2003-229341/22.

XX DR P-PSDB; ADC08177.

XX PT New plant genes encoding polypeptides having an activity involved in or

XX PS associated with the synthesis, metabolism or degradation of carbohydrates

XX PS in the plant grain useful in generating plants having improved

XX PS nutritional properties.

XX PS Claim 35; SEQ ID NO 443; 130pp; English.

XX CC This invention, in the area of plant biotechnology, relates to novel
 CC polynucleotides comprising a nucleotide sequence encoding a protein which
 CC is involved in or associated with the synthesis, metabolism or
 CC degradation of carbohydrates in the plant grain and the expression of
 CC which is up-regulated during grain filling. The plant is selected from
 CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
 CC sugarbeet, wheat, and rice. The invention may be useful for the
 CC improvement of protein, oil, starch, fibre and moisture content of the
 CC cereal grains. In addition, carbohydrate levels may be modified to a more
 CC desirable level using the present invention. The present sequence is a
 CC DNA sequence encoding a rice protein of the invention. Note: The sequence
 CC data for this patent did not form part of the printed specification, but
 CC was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/publishedpct_sequences.

XX SQ Sequence 2031 BP; 388 A; 659 C; 728 G; 246 T; 0 U; 10 Other;

XX Query Match 1.0%; Score 56.4; DB 9; Length 2031;

XX Best Local Similarity 46.7%; Pred. No. 0.00027;

Matches 214; Conservative 0; Mismatches 241; Indels 3; Gaps 1;
 QY 3348 CGGAGATTGGAACGCTTCCACCTCGGGGAGGACACAGGATAGGCCCGAGGTCAACGGCCAC 3407
 Db 942 CGCGCGTCTGGCGCGCTCGCACTCGCGCGGAGCTTCCACGCGCGCGCGCTCGCGCGC 883
 QY 3408 GGCTCTGGTGTGGGCATCACCTTTTCCCTCGAGCCAGAGGACGACGCTCCAGGGT 3467
 Db 882 GATGTCGCGCCGTCCTTCGGGCGATGACCTTGACGAAGTGTCTCATGAGGCGTTGGTC 823
 QY 3468 GGGG---CGGATTCGCTGTCCTGTAGGCTTTGACCCAGTTGAAGGAGAGCAGCGCTT 3524
 Db 822 GAAGTCTCTCGCGCGAGGTGGGTGTCGCGCTTGGTGGCAAGGACCTCGAAACACGCGCTT 763
 QY 3525 GGCCATGCTTAGGGGAGGTTCTCCAGGGGTAGATGATCGTCCAGGCGGAGCTGGC 3584
 Db 762 GTCGATGCGGAGGATGCTGACGTGCAACGTGCGCGCGGAGGTGCAAGACGAGGACGTT 703
 QY 3585 TTTCCCTGGCTCAGCCGACGCGCCAAAGGGTGCCTGATGCGCCGCGCGCTCCCGCTTGAC 3644
 Db 702 CTTCTCGCGCCCTTCTTGTGATGCGTAGGCGATGGCGGCGCGCTCGCTCGTTGAT 643
 QY 3645 GGTGTTGTGAAGACGTCGGTGGCCAGGCGCTTTTCTCAAGGACCTTCTTCAGGC 3704
 Db 642 GATCGGTGACGCTGAGCCCGCGGATGACGCGCGCTCTCTGGTGGCTTGGCGCTGCGC 583
 QY 3705 GTGACGCTCTCGCGGTGACCCCGAGTTGAGGGCCAAACATCTCCAGGGGAGCATGAA 3764
 Db 582 GTCGTTGAAGTAGGCGGAGCGGTGACGCGCGCGGTGACCTTCTCGCGAGGTAGGC 523
 QY 3765 GAGACCGTCCCGACCTTCTCCAGAGCTCCCGGTTC 3802
 Db 522 CTCGGCGTCTCTTCTATCCGGTGGACCATGGCGC 485

RESULT 12

ADA48615/c
 ID ADA48615 standard; DNA; 2028 BP.

XX AC ADA48615;

XX DT 20-NOV-2003 (first entry)

XX DE Rice gene conferring disease resistance in plants.

XX KW disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant.

XX OS Oryza sativa.

XX PN WO200300906-A2.

XX PD 03-JAN-2003.

XX PF 21-JUN-2002; 2002WO-1B002453.

XX PR 22-JUN-2001; 2001US-0300112P.

XX PR 26-SEP-2001; 2001US-0352277P.

XX PR 22-MAR-2002; 2002US-0366535P.

XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PI Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;

XX PI Katagiri F, Kreps J, Provart N, Rieke D, Zhu T;

XX DR WPI; 2003-184052/18.

XX DR P-PSDB; ADA48616.

XX PT New polynucleotide comprising a plant nucleotide sequence having an open
 PT reading frame that encodes a polypeptide associated with disease
 PT resistance, useful for conferring resistance or tolerance to a plant
 PT pathogen.

XX PS Claim 1; SEQ ID NO 685; 299pp; English.

XX The invention relates to a novel isolated polynucleotide comprising a
CC plant nucleotide sequence having an open reading frame that encodes a
CC polypeptide associated with disease resistance or its fragment having
CC substantially the same activity as the full-length polypeptide. The
CC polynucleotide of the invention is useful for conferring resistance or
CC tolerance to a plant pathogen. The present sequence represents a gene
CC conferring disease resistance used in the invention.
XX
SQ Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;

Query Match 0.9%; Score 55.4; DB 9; Length 2028;
Best Local Similarity 46.5%; Pred. No. 0.00051;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

QY 3348 CGGGAGTTTGAACGCTCCACCTCGGGAGGACAGGATGAGGCGGAGGTCACGGCCAC 3407
Db 942 CGCGCGCTTGGCGGCTCGCACTCGCGGAGGTTGCCAGCGCGCGGTCGCGGC 883

QY 3408 GGTCTTGTTGGTGGGATACCCCTTTCCCTGAGCCAGAGACACGACGTCAGGGT 3467
Db 882 GATGTGCGCGCGCTGCTTCCGCGGATGACCTTGACGAAGTGTGTCATGAGGCGTTGGTC 823

QY 3468 GGGG---CGGATTCCGTGCTCTGTTAGGCTTGCACCCAGTTGAAGGAGACGCGCGTT 3524
Db 822 GAATGCTCTCGCGCGAGGTGGGTGTCGCCCTTGTGCGAAGACCTCGAACACGCCGTT 763

QY 3525 GGCATGTCTAGGGCGAGGTTCTCCAGGGGTAGATGTAGTCGTCAGGGTGAACCTGGC 3584
Db 762 GTCGATGCGGAGGATGCTGACGTCGAACTGCGCGCGCGGAGGTGGAAGACGAGGCTT 703

QY 3645 GGTCTTGTTGAAGGACGTCGTTGCGCCACAGGCGCTTTTCTCAAGACCTTCTCCAGGC 3704
Db 642 GATGCGGTCGACGCTGAGCGCGGATGACGCGCGGCTCTTGTGGCTTCGCGCTGCGC 583

QY 3705 GTGAGCGTCTGCGGCTGACCCCGAGTTGAGGCGCCACATCTCCAGGGGACCATGAA 3764
Db 582 GTCGTTGAAGTAGGCGCGGACGCTGACGACGCGCGGCTGACCTTCTCGCGAGGTAGGC 523

QY 3765 GACGACCTCTCCACCTTCTCCAGAGCTCCCGGTTGC 3802
Db 522 CTCGCGCGCTCTCTTCACTCGGCTGAGCACCATGGCGC 485

RESULT 13
ADC08212/c
ID ADC08212 standard; DNA; 2028 BP.
XX
AC ADC08212;
XX
DT 18-DEC-2003 (first entry)
XX
DE Rice DNA sequence Seq ID517 related to grain filling.
XX
KW plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
KW gene; ds; plant.
XX
OS Oryza sativa.
XX
FN WO2003000905-A2.
XX
PD 03-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-IB002450.
XX
PR 22-JUN-2001; 2001US-0300112P.

PR 26-SEP-2001; 2001US-0325277P.
PR 20-DEC-2001; 2001US-0342327P.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;
XX
DR WPI; 2003-229341/22.
DR P-PSDB; ADC08213.
XX
PT New plant genes encoding polypeptides having an activity involved in or
PT associated with the synthesis, metabolism or degradation of carbohydrates
PT in the plant grain useful in generating plants having improved
PT nutritional properties.
XX
PS Claim 35; SEQ ID NO 517; 130pp; English.
XX
CC This invention, in the area of plant biotechnology, relates to novel
CC polynucleotides comprising a nucleotide sequence encoding a protein which
CC is involved in or associated with the synthesis, metabolism or
CC degradation of carbohydrates in the plant grain and the expression of
CC which is up-regulated during grain filling. The plant is selected from
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC sugarbeet, wheat, and rice. The invention may be useful for the
CC improvement of protein, oil, starch, fibre and moisture content of the
CC cereal grains. In addition, carbohydrate levels may be modified to a more
CC desirable level using the present invention. The present sequence is a
CC DNA sequence encoding a rice protein of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/publishedpct_sequences.
XX
SQ Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;

Query Match 0.9%; Score 55.4; DB 9; Length 2028;
Best Local Similarity 46.5%; Pred. No. 0.00051;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

QY 3348 CGGGAGTTTGAACGCTCCACCTCGGGAGGACAGGATGAGGCGGAGGTCACGGCCAC 3407
Db 942 CGCGCGCTTGGCGGCTCGCACTCGCGGAGGTTGCCAGCGCGCGGTCGCGGC 883

QY 3408 GGTCTTGTTGGGATACCCCTTTCCCTGAGCCAGAGACACGACCTCCAGGT 3467
Db 882 GATGTGCGCGCGCTGCTTCCGCGGATGACCTGACGAAGTGTGTCATGAGGCTTGGTC 823

QY 3468 GGGG---CGGATTCCGTGCTCTGTTAGGCTTGCACCCAGTTGAAGGAGACGCGCGTT 3524
Db 822 GAATGCTCTCGCGCGAGGTGGGTGTCGCCCTTGTGCGAAGACCTCGAACACGCCGTT 763

QY 3525 GGCATGTCTAGGGCGAGGTTCTCCAGGGGTAGATGTAGTCGTCAGGGTGAACCTGGC 3584
Db 762 GTCGATGCGGAGGATGCTGACGTCGAACTGCGCGCGCGGAGGTGGAAGACGAGGCTT 703

QY 3585 TTTCCCTGCGCTCAGCGGAGCGCCAAAGGGTGTGCGATGCGCGCGGCTCTCCCGTTGAC 3644
Db 702 CTTCTCNGCGCGCTTCTTGTGATGCGGTAGGCGGATGCGCGGCGGCTCGGCTCTGAT 643

QY 3645 GGTCTTGTTGAAGGACGTCGTTGCGCCACAGGCGCTTTTCTCAAGACCTTCTCCAGGC 3704
Db 642 GATGCGGTCGACGCTGAGCGCGGATGACGCGCGGCTCTTGTGGCTTCGCGCTGCGC 583

QY 3705 GTGAGCGTCTGCGGCTGACCCCGGATGACGCGCGGCTCTTGTGGCTTCGCGCTGCGC 3764
Db 582 GTCGTTGAAGTAGGCGCGGACGCTGACGACGCGCGGCTGACCTTCTCGCGAGGTAGGC 523

QY 3765 GACGACCTCTCCACCTTCTCCAGAGCTCCCGGTTGC 3802
Db 522 CTCGCGCGCTCTCTTCACTCGGCTGAGCACCATGGCGC 485

RESULT 14

ADA70341/c
 ID ADA70341 standard; DNA; 2028 BP.
 XX AC ADA70341;
 XX XX 20-NOV-2003 (first entry)
 XX DE Rice gene, SEQ ID 3664.
 XX KW Plant; bacterial infection; fungal infection; viral infection; rice;
 XX KW gene; ds.
 XX OS Oryza sativa.
 XX PN WO2003000898-A1.
 XX PD 03-JAN-2003.
 XX XX 22-JUN-2001; 2001WO-IB001105.
 XX PF 22-JUN-2001; 2001WO-IB001105.
 XX XX (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
 XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
 XX DR WPI; 2003-175290/17.
 XX XX Identifying at least one gene involved in plant resistance or response to
 XX PT pathogenic infection for conferring resistance or tolerance to a plant to
 XX PT bacterial, fungal or viral infection by determining or detecting plant
 XX PT gene expression.
 XX XX Claim 6; SEQ ID NO 3664; 899pp; English.
 XX PA The present invention relates to a method (M1) for identifying genes
 XX CC involved in plant resistance or response to pathogenic infection. M1
 XX CC comprises identifying a gene whose expression is significantly altered in
 XX CC the incompatible interaction of plant gene expression relative to
 XX CC expression of the gene in an uninfected plant, in a mutant plant that
 XX CC does not express a gene associated with response to pathogenic infection,
 XX CC or in a corresponding incompatible or compatible interaction. (M1) is
 XX CC useful for conferring resistance to resistance or tolerance to a plant to
 XX CC bacterial, fungal or viral infection. The present sequence was used to
 XX CC illustrate the invention.
 XX XX Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;
 XX PS
 XX PT
 XX CC Query Match 0.9%; Score 54.8; DB 7; Length 2028;
 XX CC Best Local Similarity 46.5%; Pred. No. 0.00074;
 XX CC Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;
 QY 3348 CGGAGTTTGGACCGTCCACCTCGGGAGGACGAGTACGAGCCAGGTCAACGCCAC 3407
 Db 942 CGGCGCTTGGCGGCTCGCACTCGCGGGAGCTTGCCTCAGCGCGCTGCCCGC 883
 QY 3408 GGTCTTGGTGTGGGATCACCTTTCCCTGAGCCAGAGACGAGCAGCTCCAGGT 3467
 Db 882 GATGTCGCGCGCTGTCTCGGGGATGACCTTACGAGTGTCTCAGCGCTTGTGC 823
 QY 3468 GGGG---CGGATTCCTGTCTCTGTAGCCCTTGACCCAGTTGAAGAGAGACGCCCTT 3524
 Db 822 GAAGTCTCTGCGCGCGAGGTGGGTGTCGCGCTTGGTGGCAAGACCTCGAACACGCCGT 763
 QY 3525 GGCATGTCTAGGCGAGTTCTCTCAGGGGTAGATGTCTCAGGGTGAGCTCGAGGTGAGCTTCG 3584
 Db 762 GTGATGGCGAGGATCTGACGTGMACTGTCGCGCGCGCGAGGTCTGAAGACGAGACGTT 703
 QY 3585 TTTCCCTGCGCTCAGCGGACGCGCCAAAGGGTTCGATGGCGCGCTCCCTCCCTTGC 3644
 Db 702 CTCTCAGCGCGCTTCTTGTGATGCGGTAGGCGATGGCGGCGGCGGTCTGCTGAT 643

QY 3645 GGTTTGGTGAAGGACGTCGGTGGCCACAGGCGCTTTTCTCAAGGACCTTCTTCAGGC 3704
 Db 642 GATCGGTCGACGGTGAGCCCGGATGACGCGGGCTCTTGGTGGCTTGGCGCTGCCG 583
 QY 3705 GTGACGGTCTGCCGGTGACCCCGAGTTGAGGGCCAACTCTCCAGGGGACCATGAA 3764
 Db 582 GTCGTTGAAGTAGGCCCGGACCGGTGACGACGGCGCGTACCTTCTCGCCGAGGTAGC 523
 QY 3765 GACGACCGTCCCGACCTTCTCCAGAGCTCCCGGTTGC 3802
 Db 522 CTCGGCGTCTCTCTTCATCCGGTGAGCACCATGGCGC 485
 RESULT 15
 ADA71232/c
 ID ADA71232 standard; DNA; 2060 BP.
 XX AC ADA71232;
 XX XX 20-NOV-2003 (first entry)
 XX DE Rice gene, SEQ ID 4555.
 XX KW Plant; bacterial infection; fungal infection; viral infection; rice;
 XX KW gene; ds.
 XX OS Oryza sativa.
 XX PN WO2003000898-A1.
 XX PD 03-JAN-2003.
 XX PF 22-JUN-2001; 2001WO-IB001105.
 XX XX 22-JUN-2001; 2001WO-IB001105.
 XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
 XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
 XX DR WPI; 2003-175290/17.
 XX XX Identifying at least one gene involved in plant resistance or response to
 XX PT pathogenic infection for conferring resistance or tolerance to a plant to
 XX PT bacterial, fungal or viral infection by determining or detecting plant
 XX PT gene expression.
 XX XX Claim 6; SEQ ID NO 4555; 899pp; English.
 XX PS The present invention relates to a method (M1) for identifying genes
 XX CC involved in plant resistance or response to pathogenic infection. M1
 XX CC comprises identifying a gene whose expression is significantly altered in
 XX CC the incompatible interaction of plant gene expression relative to
 XX CC expression of the gene in an uninfected plant, in a mutant plant that
 XX CC does not express a gene associated with response to pathogenic infection,
 XX CC or in a corresponding incompatible or compatible interaction. (M1) is
 XX CC useful for conferring resistance to resistance or tolerance to a plant to
 XX CC bacterial, fungal or viral infection. The present sequence was used to
 XX CC illustrate the invention.
 XX XX Sequence 2060 BP; 395 A; 670 C; 739 G; 254 T; 0 U; 2 Other;
 XX PS
 XX PT
 XX CC Query Match 0.9%; Score 54.8; DB 7; Length 2060;
 XX CC Best Local Similarity 46.5%; Pred. No. 0.00075;
 XX CC Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;
 QY 3348 CGGAGTTTGGACCGTCCACCTCGGGAGGACGAGTACGAGCCAGGTCAACGCCAC 3407
 Db 942 CGGCGCTTGGCGGCTCGCACTCGCGGGAGCTTGCCTCAGCGCGCTGCCCGC 883
 QY 3408 GGTCTTGGTGTGGGATCACCTTTCCCTGAGCCAGAGACGAGCAGCTCCAGGT 3467

Db 882 GATGTCGGCCGCTGCTTCGGGGGATGACCTTGACGAAGTGGTCCATGAGGCGTTGGTC 823
Qy 3468 GGGG---CGGATTCCTGGTCTCTGCTAGGCTTTGACCCAGTTGAAGGAGAGACGCGTT 3524
Db 822 GAAGTCTCTCGCCCGCGAGGTGGGTCTGCGCGTTGGTGGCAAGGACCTTCGAACACGCGTT 763
Qy 3525 GGCCATGTCTAGGGCGAGGTTCTCTCCAGGGGTAGATAGTCTGCTCCAGGGTGAGCCTGGC 3584
Db 762 GTCGATGGCGAGGATGCTGACGTGCAACGTGCGCCGCCGAGGTCTGAAGACGAGGACGTT 703
Qy 3585 TTTCCCTGGCTCAGCCGAGCGGCCAAAGGGTCCGATGGCCCGGCGCTCCCGTTGAC 3644
Db 702 CTTCCTCGCGCCCTTCTTGTGATGCCGTAGGCGATGGCGCGCGGTCGECTCGTTGAT 643
Qy 3645 GGTTTGGTGAAGGACGTGCGGTGGCCACGAGGCCCTTTTCTCAAGGACCTTCTTCCAGGC 3704
Db 642 GATCGGTCGACGTGAGCCCGGCGATGACGCCGGCGTCTTGGTGGCTTCCGCTCGC 583
Qy 3705 GTGACCGTCTGCGGGTGACCCCGAGGTTGAGGGCCAAACATCTCCAGGGGGACCATGAA 3764
Db 582 GTCGTTGAAGTAGGCCCGGACGCTGACGACGGCGCGGTGACCTTCTCGCCGAGGTAGGC 523
Qy 3765 GACGACGTCCTCCACCTTCTCCAGAGTCCCGGTTGC 3802
Db 522 CTCGGCGCTCTCTTCATCCGGGTGAGCACCATGGCGC 485

Search completed: March 17, 2004, 21:48:03
Job time : 1433.28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 21:15:22 ; Search time 273.095 Seconds
(without alignments)
11885.630 Million cell updates/sec

Title: US-09-664-186-6
Perfect score: 5849
Sequence: 1 tctagaaggtcagggtggac.....ttagaggaccagaagacc 5849

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgm2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgm2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgm2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgm2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgm2_6/ptodata/2/ina/6CTUS COMB.seq.*
6: /cgm2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	5849	100.0	5849	3	US-09-134-246-6
2	1242	21.2	1242	2	US-08-960-756-1
3	1026	17.5	1026	3	US-09-134-246-4
4	999	17.1	999	2	US-08-960-756-3
5	180	3.1	180	3	US-09-134-246-5
6	77	1.3	7218	1	US-08-232-463-14
7	56.4	1.0	7218	1	US-08-232-463-14
8	51.8	0.9	1551	4	US-09-252-991A-5309
9	51.8	0.9	1563	4	US-09-252-991A-15934
10	51.8	0.9	1953	4	US-09-252-991A-15763
11	51.8	0.9	3351	4	US-09-252-991A-15871
12	51.8	0.9	4158	4	US-09-252-991A-5348
13	51.8	0.9	4953	4	US-09-252-991A-5227
14	49.6	0.8	4403765	3	US-09-103-840A-2
15	49.6	0.8	4411529	3	US-09-103-840A-1
16	46.8	0.8	2460	4	US-09-252-991A-2998
17	46.8	0.8	2856	4	US-09-252-991A-2869
18	46.8	0.8	3387	4	US-09-252-991A-3101
19	46.2	0.8	1473	4	US-09-152-060-43
20	46	0.8	1572	4	US-09-252-991A-8725
21	46	0.8	1929	4	US-09-252-991A-8827
22	46	0.8	2079	4	US-09-252-991A-9222
23	45.8	0.8	960	4	US-09-252-991A-6990
24	45.8	0.8	1276	4	US-09-252-991A-7059
25	45.8	0.8	1925	4	US-09-249-585A-4
26	45.8	0.8	1931	2	US-09-130-114-2
27	45.8	0.8	2187	4	US-09-252-991A-7042

ALIGNMENTS

RESULT 1

US-09-134-246-6

; Sequence 6, Application US/09134246B

; Patent No. 6207377

; GENERAL INFORMATION:

; APPLICANT: Wayne, Jay

; APPLICANT: Xu, Shuang-yong

; TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle

; TITLE OF INVENTION: Vectors And Identification Of Two Thermus Plasmid

; TITLE OF INVENTION: Replication Origins

; FILE REFERENCE: Thermus Shuttle Vector

; CURRENT APPLICATION NUMBER: US/09/134,246B

; CURRENT FILING DATE: 1998-08-14

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 6

; LENGTH: 5849

; TYPE: DNA

; ORGANISM: Thermus sp.

US-09-134-246-6

Query Match	100.0%;	Score 5849;	DB 3;	Length 5849;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 5849;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	TCTAGAGGTGAGGTGGACAGGAAACACCATAGCCCTCCAGAGATGGAGCT	60	
DB	1	TCTAGAGGTGAGGTGGACAGGAAACACCATAGCCCTCCAGAGATGGAGCT	60	
QY	61	TGCTGTCGGGAAAGTGGCCATCCGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA	120	
DB	61	TGCTGTCGGGAAAGTGGCCATCCGGGGCGCTCTTGACAACTATTTTCCAGCGGTGGCCA	120	
QY	121	CCGGCATTTGGCCACGAGGTGACAGCTTTGTGGAGTAGACGCCACAAAGGGTGTCTCA	180	
DB	121	CCGGCATTTGGCCACGAGGTGACAGCTTTGTGGAGTAGACGCCACAAAGGGTGTCTCA	180	
QY	181	AACCTCTTTTCTAGTCCCGCTTGAGCAAGGGGAGAGAGAAAGCTTCATGGCCTCA	240	
DB	181	AACCTCTTTTCTAGTCCCGCTTGAGCAAGGGGAGAGAGAAAGCTTCATGGCCTCA	240	
QY	241	CCTCTTCCCTCTCTCTTTGGGGCGCTTAGCGCGTAAACTCTGAGACGGCTGAAGTT	300	
DB	241	CCTCTTCCCTCTCTCTTTGGGGCGCTTAGCGCGTAAACTCTGAGACGGCTGAAGTT	300	
QY	301	TAGGGATTCGCTTTTCGGGATTAAGATCCGGCGGTACGGGATCCCGATGGCCCTTA	360	
DB	301	TAGGGATTCGCTTTTCGGGATTAAGATCCGGCGGTACGGGATCCCGATGGCCCTTA	360	
QY	361	TCCTGCCGCTCTTATGTACTCGTAAATGGTGGCTTGGTACTTTTAAACCGTCTTGA	420	

Db 361 |||||TCCTGCGTCCCTTATGTACTCGTAATAGTGGCCCTTGGGTACTTTAAACCGTTCGAAA 420
Qy 421 CTTTCTTAACAGAGACACAAAACCTCTAAAACCTATCAATCCACCGATCCAGTATA 480
Db 421 CTTTCTTAACAGAGACACAAAACCTCTAAAACCTATCAATCCACCGATCCAGTATA 480
Qy 481 CCAATAATGGCACAAGTTTGGAGAGTGGTCAACAAAAGGCTTTCTCGGTCAAGTT 540
Db 481 CCAATAATGGCACAAGTTTGGAGAGTGGTCAACAAAAGGCTTTCTCGGTCAAGTT 540
Qy 541 ATGCTGAGGTGGGGCGGTCAAAGGCGGACTTAAAGTTTGGTAAAGCCGGAGGAAAGCAAA 600
Db 541 ATGCTGAGGTGGGGCGGTCAAAGGCGGACTTAAAGTTTGGTAAAGCCGGAGGAAAGCAAA 600
Qy 601 CCGGGTGTATCCATGCAACAGATGGCCGAGTGGAAACGTTGGTAAAGCCGGAGGAAAGCAAA 660
Db 601 CCGGGTGTATCCATGCAACAGATGGCCGAGTGGAAACGTTGGTAAAGCCGGAGGAAAGCAAA 660
Qy 661 GCTTCTGGAGAGGGGTATTTGGATAAACTACTGAGGTCTATAAGGGGAAAGTGGCTC 720
Db 661 GCTTCTGGAGAGGGGTATTTGGATAAACTACTGAGGTCTATAAGGGGAAAGTGGCTC 720
Qy 721 TTCGAGTCACTACAGAGAGGTAGAGAAAACCTTCGCGAGGCTTACAGGCAATACGA 780
Db 721 TTCGAGTCACTACAGAGAGGTAGAGAAAACCTTCGCGAGGCTTACAGGCAATACGA 780
Qy 781 GGGAGGAGGATAGTCCGAGGCGAAGACGAACTCGTGGAGCGGCTTAATGCCAG 840
Db 781 GGGAGGAGGATAGTCCGAGGCGAAGACGAACTCGTGGAGCGGCTTAATGCCAG 840
Qy 841 AAAAAAGTTCGAGCGGTCCCGCTTCAATCACCCCTACCTGCTTGGTCTACTACCTGGT 900
Db 841 AAAAAAGTTCGAGCGGTCCCGCTTCAATCACCCCTACCTGCTTGGTCTACTACCTGGT 900
Qy 901 TTCGAAAAAGCAGAAACGAAACGAACTCGAGGCGCTTGGAGGCTTCCAGGAGTTCCTC 960
Db 901 TTCGAAAAAGCAGAAACGAAACGAACTCGAGGCGCTTGGAGGCTTCCAGGAGTTCCTC 960
Qy 961 AAAGCACCAGAAACCATCCGGTCTCGGCAAGGAGCGCAAGAAAGGCGGTAGAAGC 1020
Db 961 AAAGCACCAGAAACCATCCGGTCTCGGCAAGGAGCGCAAGAAAGGCGGTAGAAGC 1020
Qy 1021 CTTGATCCAAAGGCTCAAGAGCTCCCGAATAAATCGGCAAGATAGGCGCGATGTTCAA 1080
Db 1021 CTTGATCCAAAGGCTCAAGAGCTCCCGAATAAATCGGCAAGATAGGCGCGATGTTCAA 1080
Qy 1081 AAGGTGTCACAAAGAGCTAAAGGGGAAATAGAGAGGCTTCCAGGCGCTTACCAA 1140
Db 1081 AAGGTGTCACAAAGAGCTAAAGGGGAAATAGAGAGGCTTCCAGGCGCTTACCAA 1140
Qy 1141 ACCAAAGATTGTAGTATCCCTGAAAAAAGTAAACCGGAGCAAGCAACCCCTATTTC 1200
Db 1141 ACCAAAGATTGTAGTATCCCTGAAAAAAGTAAACCGGAGCAAGCAACCCCTATTTC 1200
Qy 1201 GGAGAGAGCGGGCATCATCATATACCGGATCGGATGAGCTTTCAAGATGCGCGC 1260
Db 1201 GGAGAGAGCGGGCATCATCATATACCGGATCGGATGAGCTTTCAAGATGCGCGC 1260
Qy 1261 CAAAGAAAACTGGGCGCTTGGGAGGAGCAAGAACTAGGACCAAGAGGCGTATGTTCTA 1320
Db 1261 CAAAGAAAACTGGGCGCTTGGGAGGAGCAAGAACTAGGACCAAGAGGCGTATGTTCTA 1320
Qy 1321 CGTGTATATCCCGCTAGCGCTTAAGAGACATGGCACCTTAACAGGAGAGTGAAGTTTCA 1380
Db 1321 CGTGTATATCCCGCTAGCGCTTAAGAGACATGGCACCTTAACAGGAGAGTGAAGTTTCA 1380
Qy 1381 ATCCGACTTTGGCGGAAAACCAAGACAAACAGAAAACCTAGTAGCAAAAGGCTTCCATAAGGT 1440
Db 1381 ATCCGACTTTGGCGGAAAACCAAGACAAACAGAAAACCTAGTAGCAAAAGGCTTCCATAAGGT 1440
Qy 1441 GGAACCTTGAGAGAGGCAATAGGAATAGTGGTGGAGCGGAATGCCTGGGTGAGCAA 1500

Db 1441 GGAACCTTGAGAGAGGCAATAGGAATAGTGGTGGAGCGGAATGCCTGGGTGAGCAA 1500
Qy 1501 GTTTCGTGGGTGGGCGGACTGGGGAAGAAACGATCGTTACATCGGTACTCCTCTTCC 1560
Db 1501 GTTTCGTGGGTGGGCGGACTGGGGAAGAAACGATCGTTACATCGGTACTCCTCTTCC 1560
Qy 1561 AGACCTGATAGCGAGCTCTACCAAAAGGGAAGAACCCCTGGGCTCTAGAAAGCGGA 1620
Db 1561 AGACCTGATAGCGAGCTCTACCAAAAGGGAAGAACCCCTGGGCTCTAGAAAGCGGA 1620
Qy 1621 CACAATCTCAAACTTGTGTAGCCCTGGGAAATCTCTAACACCTTCTAGTGAAGGC 1680
Db 1621 CACAATCTCAAACTTGTGTAGCCCTGGGAAATCTCTAACACCTTCTAGTGAAGGC 1680
Qy 1681 TTTGACCGCTCCAGAGAGCATCTATGCCATGGATCGCGCTTTAAGAGGGGTGAGGC 1740
Db 1681 TTTGACCGCTCCAGAGAGCATCTATGCCATGGATCGCGCTTTAAGAGGGGTGAGGC 1740
Qy 1741 TATAAGCGTAGTACCGAGAGCTTCGAAAGGATCGAGCACTAAATCCCTCTGTTACTCC 1800
Db 1741 TATAAGCGTAGTACCGAGAGCTTCGAAAGGATCGAGCACTAAATCCCTCTGTTACTCC 1800
Qy 1801 TGTTCGACCATGAGCTTGAGATGTCAGATTTTCTCGGTGGGTATCGCGGTACCG 1860
Db 1801 TGTTCGACCATGAGCTTGAGATGTCAGATTTTCTCGGTGGGTATCGCGGTACCG 1860
Qy 1861 AGGATCCTTGAACTGCGCAACGCTCTGGAGCTTTTCCCTCTTTCAGGCGATCCCGAGC 1920
Db 1861 AGGATCCTTGAACTGCGCAACGCTCTGGAGCTTTTCCCTCTTTCAGGCGATCCCGAGC 1920
Qy 1921 GTAAACTTTCTTCGCGGCAACCCGCTTTGACCAACAATAAGCCCTTGACGCTCTAG 1980
Db 1921 GTAAACTTTCTTCGCGGCAACCCGCTTTGACCAACAATAAGCCCTTGACGCTCTAG 1980
Qy 1981 CTGCTCAAGCTTCTTCGCGGCAATAGCCCAATCCCGCTTCAGAGAGGGGAAGTATTCCTCG 2040
Db 1981 CTGCTCAAGCTTCTTCGCGGCAATAGCCCAATCCCGCTTCAGAGAGGGGAAGTATTCCTCG 2040
Qy 2041 CCAAGGCTTCGCTAGGCGCATCTTGTGTTCTCCAGAGCATGCGAGGATGTTGGT 2100
Db 2041 CCAAGGCTTCGCTAGGCGCATCTTGTGTTCTCCAGAGCATGCGAGGATGTTGGT 2100
Qy 2101 GTACCGTTCCCGCTTCGCTACAAAGGGAAAGCCTAGCGATCTCTTTCGAAATA 2160
Db 2101 GTACCGTTCCCGCTTCGCTACAAAGGGAAAGCCTAGCGATCTCTTTCGAAATA 2160
Qy 2161 GGGCTAGCGGATTCGTTCCAAACGTAAGTCCCGCTTTGGAGTAGACGAGGATCATGTC 2220
Db 2161 GGGCTAGCGGATTCGTTCCAAACGTAAGTCCCGCTTTGGAGTAGACGAGGATCATGTC 2220
Qy 2221 CTTTTCGATCCGAGGCTTACGGGAAAGTTTTGGGATTTGAAGCGATGCGGCGAT 2280
Db 2221 CTTTTCGATCCGAGGCTTACGGGAAAGTTTTGGGATTTGAAGCGATGCGGCGAT 2280
Qy 2281 ATGCTTAACGAAGTTTCGCGGCAAGACCTCATCAAGGATGAGCTTCACTCGAACCC 2340
Db 2281 ATGCTTAACGAAGTTTCGCGGCAAGACCTCATCAAGGATGAGCTTCACTCGAACCC 2340
Qy 2341 GTATTTCTGCTATGTGAACGAAGATCAGTCTGAGTCCGCGATCAGTCTCCCTGAGAAG 2400
Db 2341 GTATTTCTGCTATGTGAACGAAGATCAGTCTGAGTCCGCGATCAGTCTCCCTGAGAAG 2400
Qy 2401 TATCAAGCGCTCCCTCAGGAACTCCAAACTCAGGAGCATCGAGGCTGTCATCGTAGCC 2460
Db 2401 TATCAAGCGCTCCCTCAGGAACTCCAAACTCAGGAGCATCGAGGCTGTCATCGTAGCC 2460
Qy 2461 CAACTGACCGTTTTCGCGTGGCTGACGAGTAAAGCGATCTGTTTCAATCGCGCGAAC 2520
Db 2461 CAACTGACCGTTTTCGCGTGGCTGACGAGTAAAGCGATCTGTTTCAATCGCGCGAAC 2520
Qy 2521 GAGAACTGCTGCGGTTCCATAAGCGGCTCAATATAGACCAACTGACCTTCCCGC 2580
Db 2521 GAGAACTGCTGCGGTTCCATAAGCGGCTCAATATAGACCAACTGACCTTCCCGC 2580

Qy 2581 ATACCCACAGGCTCCCGAGGATCCACCGGAACTGACCGTTTCCCCAAAAGTA 2640
Db 2581 ATACCCACAGGCTCCCGAGGATCCACCGGAACTGACCGTTTCCCCAAAAGTA 2640
Qy 2641 GGTGCCAATAGGATCAATCTCAAAAGGGGGGCAATTTCCCCCTAGGAAGAGAGGTTTC 2700
Db 2641 GGTGCCAATAGGATCAATCTCAAAAGGGGGGCAATTTCCCCCTAGGAAGAGAGGTTTC 2700
Qy 2701 TTTTCGAAAACAAGTTGTGGGTGGCTGATCAAGAACTCTCTTCTCATCGGTTTCC 2760
Db 2701 TTTTCGAAAACAAGTTGTGGGTGGCTGATCAAGAACTCTCTTCTCATCGGTTTTC 2760
Qy 2761 GGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGCTTT 2820
Db 2761 GGGTAGACCAACCTAAAGGGCGAAGGTTCCGAGGTTTTCGAGGCTTTCAAGGGGCTTT 2820
Qy 2821 TCGGCTCAAAACCGAGGTAGCTACGGCTCATTTCTTCCCTCCCAAGGCTTTTAAGCAGG 2880
Db 2821 TCGGCTCAAAACCGAGGTAGCTACGGCTCATTTCTTCCCTCCCAAGGCTTTTAAGCAGG 2880
Qy 2881 ACCTCATCACCAACACCTCAGCACTCCAAACCAAGGAATCCGCAAGGGCGCTACC 2940
Db 2881 ACCTCATCACCAACACCTCAGCACTCCAAACCAAGGAATCCGCAAGGGCGCTACC 2940
Qy 2941 TTTTAGCCCGGTATCTTCCCTGACGTPATAGACTTCGGATCGTCTCAGGGTCAACCGGA 3000
Db 2941 TTTTAGCCCGGTATCTTCCCTGACGTPATAGACTTCGGATCGTCTCAGGGTCAACCGGA 3000
Qy 3001 AGGATGTCTCAAGCTCTCGGGGTGAGTACACGGGCTTCTCCTCATGACACAACT 3060
Db 3001 AGGATGTCTCAAGCTCTCGGGGTGAGTACACGGGCTTCTCCTCATGACACAACT 3060
Qy 3061 TACCCCAAGAGGACAAACATGCAACTATGGGCAAAAGTAGACAAAGAGCTT 3120
Db 3061 TACCCCAAGAGGACAAACATGCAACTATGGGCAAAAGTAGACAAAGAGCTT 3120
Qy 3121 GGGCACTCTCTCAGAGGCTCTTGAAGGCTTTTCACTAGGTACCGTCCCGTCTGTC 3180
Db 3121 GGGCACTCTCTCAGAGGCTCTTGAAGGCTTTTCACTAGGTACCGTCCCGTCTGTC 3180
Qy 3181 AGATGGCCATCCGTGTAATCTCGGATAACCGCAATAGGACGGCAATAGATTTGCGG 3240
Db 3181 AGATGGCCATCCGTGTAATCTCGGATAACCGCAATAGGACGGCAATAGATTTGCGG 3240
Qy 3241 GGGAGTTCAACCCCTGGCCACAGCCACAGCAAGCTGCATAGAAACGTCTTGAACGACGG 3300
Db 3241 GGGAGTTCAACCCCTGGCCACAGCCACAGCAAGCTGCATAGAAACGTCTTGAACGACGG 3300
Qy 3301 TCATCTAGGATCGGCAATGTAGTAGCAATGAGGTTGATAAGGCGGGAGTTTGGAA 3360
Db 3301 TCATCTAGGATCGGCAATGTAGTAGCAATGAGGTTGATAAGGCGGGAGTTTGGAA 3360
Qy 3361 CGCTCCACCTCGGGAGGACCAAGGATGAGGCCCAAGGTCAAGGCCACGGTCTTGGTGTG 3420
Db 3361 CGCTCCACCTCGGGAGGACCAAGGATGAGGCCCAAGGTCAAGGCCACGGTCTTGGTGTG 3420
Qy 3421 GGCATACCCCTTTCCCTGAGCCAGAGGACACAGCAGCTCCAGGTTGGGCGGATTCG 3480
Db 3421 GGCATACCCCTTTCCCTGAGCCAGAGGACACAGCAGCTCCAGGTTGGGCGGATTCG 3480
Qy 3481 TGGTCTGGTAGGCTTGACCCAGTTGAAGAGACACCGCTTGGCCATGCTTAGGGCG 3540
Db 3481 TGGTCTGGTAGGCTTGACCCAGTTGAAGAGACACCGCTTGGCCATGCTTAGGGCG 3540
Qy 3541 AGGTTCTTCAGGGGTAGATGTGCTCAGGGGTAGCTTGGCTTTCCCTGGCTCAGC 3600
Db 3541 AGGTTCTTCAGGGGTAGATGTGCTCAGGGGTAGCTTGGCTTTCCCTGGCTCAGC 3600
Qy 3601 CGGAGGCCCCAAGGTTCCGATGCGTCCCGGCTCCCGTTGACGTTTGGTGAAGACG 3660
Db 3601 CGGAGGCCCCAAGGTTCCGATGCGTCCCGGCTCCCGTTGACGTTTGGTGAAGACG 3660

Qy 3661 TCGGTGGCCACACGAGCCCTTTTCTCAAGGACCTTTCTCAGGCGTGAACGCTTCGCGG 3720
Db 3661 TCGGTGGCCACACGAGCCCTTTTCTCAAGGACCTTTCTCAGGCGTGAACGCTTCGCGG 3720
Qy 3721 GTGACCCCAAGGTTGAGGCCCCAAACATCTCAGGGGGACCATGAAGACGACGTCGCCAC 3780
Db 3721 GTGACCCCAAGGTTGAGGCCCCAAACATCTCAGGGGGACCATGAAGACGACGTCGCCAC 3780
Qy 3781 TTTCTCCACAGAGCTCCCGGTTGCGGTAGGGGATGTTGAGCGGGCAATCTCTCGAGGAGT 3840
Db 3781 TTTCTCCACAGAGCTCCCGGTTGCGGTAGGGGATGTTGAGCGGGCAATCTCTCGAGGAGT 3840
Qy 3841 TCCAGAACTCTGAGGCCCCGCTCTGAGCGGCTTACATATCTCCAAACGGGGGTTCA 3900
Db 3841 TCCAGAACTCTGAGGCCCCGCTCTGAGCGGCTTACATATCTCCAAACGGGGGTTCA 3900
Qy 3901 GCTGTGTGTATAGGGGGGGCCCCAACAGAGGAAAGAGCTCATGTCTTTCTCGGCT 3960
Db 3901 GCTGTGTGTATAGGGGGGGCCCCAACAGAGGAAAGAGCTCATGTCTTTCTCGGCT 3960
Qy 3961 AAAGCAATCTTAAGTGCCTCTTTTGGTATGTAAAGCCCTTCGAGGCGATTTTCGGCA 4020
Db 3961 AAAGCAATCTTAAGTGCCTCTTTTGGTATGTAAAGCCCTTCGAGGCGATTTTCGGCA 4020
Qy 4021 CTTCCATCTGAGGGGGCTCGGTGCGCAAGAGAGTCTCTGACCCCTATCTGACCCC 4080
Db 4021 CTTCCATCTGAGGGGGCTCGGTGCGCAAGAGAGTCTCTGACCCCTATCTGACCCC 4080
Qy 4081 CTAGTGGCATCGGTTGTTGTTGGTCTTCTTAAAGCTCTGTAAGCTCTTCAAGAAG 4140
Db 4081 CTAGTGGCATCGGTTGTTGTTGGTCTTCTTAAAGCTCTGTAAGCTCTTCAAGAAG 4140
Qy 4141 GTTTTGTGTTCTTCAACCTCGGACCTCTGTTGATCTGAGCCCGAGGCGTTACCTAG 4200
Db 4141 GTTTTGTGTTCTTCAACCTCGGACCTCTGTTGATCTGAGCCCGAGGCGTTACCTAG 4200
Qy 4201 GTCCTGGGGTGTATCGGGGCAACCGCTCGGTTTGGCTTTTATGGGTCAAAATAAC 4260
Db 4201 GTCCTGGGGTGTATCGGGGCAACCGCTCGGTTTGGCTTTTATGGGTCAAAATAAC 4260
Qy 4261 CTTAGCCCGAGCGGTGCAATCCCGCTTCTTAAAGGCGGTTATAGGCGCTCTGAGGAG 4320
Db 4261 CTTAGCCCGAGCGGTGCAATCCCGCTTCTTAAAGGCGGTTATAGGCGCTCTGAGGAG 4320
Qy 4321 GGGGTAGTACTTCTTACCCCTTAGGCTTGGAGAGGCTTAGGAGGCTCTCTAGGGCC 4380
Db 4321 GGGGTAGTACTTCTTACCCCTTAGGCTTGGAGAGGCTTAGGAGGCTCTCTAGGGCC 4380
Qy 4381 TCGTGGGGGTGTAGGGGTAACTCATGGCCAGGCGCGGCTCGGACTCTGGAGGAG 4440
Db 4381 TCGTGGGGGTGTAGGGGTAACTCATGGCCAGGCGCGGCTCGGACTCTGGAGGAG 4440
Qy 4441 CTTCCATAGCTACTCTGTTGGTGTGTAAGGGGTTTCACTAATGCATACGGCTAGC 4500
Db 4441 CTTCCATAGCTACTCTGTTGGTGTGTAAGGGGTTTCACTAATGCATACGGCTAGC 4500
Qy 4501 CTCGGGATCAGCGCCAAATGTTGATGAGGTTTGGTATAAAACCCCTCAGGTTTGGGCTA 4560
Db 4501 CTCGGGATCAGCGCCAAATGTTGATGAGGTTTGGTATAAAACCCCTCAGGTTTGGGCTA 4560
Qy 4561 GTTTATGTCTGTTTATGCACTTTGATCTCGGATCAACGGGCTAAACACAGTTCCTGCG 4620
Db 4561 GTTTATGTCTGTTTATGCACTTTGATCTCGGATCAACGGGCTAAACACAGTTCCTGCG 4620
Qy 4621 ACGAAGAAATCTTTCGGATCTAAGAGGGGAAAGAGGTGTAGAGGGGACGGCTTCATG 4680
Db 4621 ACGAAGAAATCTTTCGGATCTAAGAGGGGAAAGAGGTGTAGAGGGGACGGCTTCATG 4680
Qy 4681 AAAGTTGGCTCTTAGGAGGCGCTTGTAGAGGCGCTCTCGGGTTCAAAATCTTTCCCTC 4740
Db 4681 AAAGTTGGCTCTTAGGAGGCGCTTGTAGAGGCGCTCTCGGGTTCAAAATCTTTCCCTC 4740
Qy 4741 TCTCTCCAGGTTCCGAGGTTGAGGCTCTTGTCCAGGCTTGTGTAACCAAGTTTGTGACCA 4800

4741	Db	TCTCTCAGGTTTCCGAGGTTTCGAGGTCCTTGGTCCAGGTCCTTGTPACCAAGTTTTTGACCA	4801
4801	Qy	AAGTCTATTCTCGGAATATAGGGGTATCTTGTGTCATCTTCCTACGGATATCTCTGCT	4860
4801	Db	AAGTCTATTCTCGGAATATAGGGGTATCTTGTCTATCTTCCCTACGGGATATCTCTGCT	4860
4861	Qy	GTGTGAACCTGATCCCATCCCAATACATATCTCAATCTCCTTAATCTCTCTCTCCAG	4920
4861	Db	GTGTGAACCTGATCCCATCCCAATACATATCTCAATCTCCTTAATCTCTCTCTCTCCAG	4920
4921	Qy	ATCCCTAATCTCTCTCTACCTCTTCTCCTCCCAATTAAGANTGAGAGGAGAAACCC	4980
4921	Db	ATCCCTAATCTCTCTCTACCTCTTCTCCTCCCAATTAAGANTGAGAGGAGAAACCC	4980
4981	Qy	CGACCAGAACGAGCTTCTCGGGTCAGTTTCGGTAAATCTCGGCACAGTTTTTCATCGTCT	5040
4981	Db	CGACCAGAACGAGCTTCTCGGGTCAGTTTCGGTAAATCTCGGCACAGTTTTTCATCGTCT	5040
5041	Qy	AGACGAGGATTAGGGCATGAAATAATCGGCTTTGACAAATCTTTTCAAAAATACTCCC	5100
5041	Db	AGACGAGGATTAGGGCATGAAATAATCGGCTTTGACAAATCTTTTCAAAAATACTCCC	5100
5101	Qy	CGAGTTTGGGGAAGTGCCTCGGGGAGAAGATTTTGGCAGTTTAGATGTTATGCTCTAT	5160
5101	Db	CGAGTTTGGGGAAGTGCCTCGGGGAGAAGATTTTGGCAGTTTAGATGTTATGCTCTAT	5160
5161	Qy	CACGGGCCGAGGCTCCACGATAAGTTGTCTTGGCCAAAGTACCGGCCACAGTCGGGGT	5220
5161	Db	CACGGGCCGAGGCTCCACGATAAGTTGTCTTGGCCAAAGTACCGGCCACAGTCGGGGT	5220
5221	Qy	GCTCTTCAGCGTGTGTATGTTTTCACGGAAGTTCACAAGTCCCTTTTAGAGGCTTCAG	5280
5221	Db	GCTCTTCAGCGTGTGTATGTTTTCACAAGGATTCACAAGTCCCTTTTAGAGGCTTCAG	5280
5281	Qy	GTCGGGATAGTGTCTCAAGTACTCCCAGCGTTCTCGGGCCGGTGTGCGGGAGAAAGCAC	5340
5281	Db	GTCGGGATAGTGTCTCAAGTACTCCCAGCGTTCTCGGGCCGGTGTGCGGGAGAAAGCAC	5340
5341	Qy	AAAGGGTTCGGGCAAAAGTTTCATCTTTGTACTTAGGACGGATTAATTGACCTGATAA	5400
5341	Db	AAAGGGTTCGGGCAAAAGTTTCATCTTTGTACTTAGGACGGATTAATTGACCTGATAA	5400
5401	Qy	CTTCAGGGCCGTTAAGAAGGCCCTCACCTCGGACACGGTTCGAGGAGGACGTGGGGCTG	5460
5401	Db	CTTCAGGGCCGTTAAGAAGGCCCTCACCTCGGACACGGTTCGAGGAGGACGTGGGGCTG	5460
5461	Qy	GAAGAACGAAACCCCGATTTTGGGAAGTCTCCCTCAGTTTGATGATGAACGTGGGA	5520
5461	Db	GAAGAACGAAACCCCGATTTTGGGAAGTCTCCCTCAGTTTGATGATGAACGTGGGA	5520
5521	Qy	GGAAGCCGGCCAGGATGCTTTTCATCGCGCTCGAACCTCGGACACATAAAAACCTTCG	5580
5521	Db	GGAAGCCGGCCAGGATGCTTTTCATCGCGCTCGAACCTCGGACACATAAAAACCTTCG	5580
5581	Qy	TGTTTGTACGGGCAAGAGTGCTATGTATAGGTAAACCTTCGGGAGTACAAAGTCCTCAA	5640
5581	Db	TGTTTGTACGGGCAAGAGTGCTATGTATAGGTAAACCTTCGGGAGTACAAAGTCCTCAA	5640
5641	Qy	GCGCCCTTCCCAACGCTCCAAAACCTAGGTCAGGTGTTTAGTTTCTCGAAAACCT	5700
5641	Db	GCGCCCTTCCCAACGCTCCAAAACCTAGGTCAGGTGTTTAGTTTCTCGAAAACCT	5700
5701	Qy	CTAGCTTTTCAGTGGTCATTCCTCACCCCTCTAGCACGTACTCTGGAAGGTAAACCTTTCG	5760
5701	Db	CTAGCTTTTCAGTGGTCATTCCTCACCCCTCTAGCACGTACTCTGGAAGGTAAACCTTTCG	5760
5761	Qy	ACACAGCGGCCAAGTCTAGGTCCTCCAGTCCAGTTGGTCTGGGACCGGTGAGAGGGGA	5820
5761	Db	ACACAGCGGCCAAGTCTAGGTCCTCCAGTCCAGTTGGTCTGGGACCGGTGAGAGGGGA	5820
5821	Qy	GGGGCTTGGTGTAGGAGCAGAAAGACCC	5849

Db 5821 GGGCGTGTGTAGAGGACGAGACCC 5849

RESULT 2

US-08-960-756-1/c

; Sequence 1, Application US/08960756

; Patent No. 5868422

; GENERAL INFORMATION:

; APPLICANT: WAYNE, JAY

; APPLICANT: XU, SHUANG-YONG

; TITLE OF INVENTION: METHOD FOR CLONING AND

; TITLE OF INVENTION: PRODUCING THE Tsp45I RESTRICTION ENDONUCLEASE IN E. COLI

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: New England Biolabs, Inc.

; STREET: 32 Tozer Road

; CITY: Beverly

; STATE: MA

; COUNTRY: US

; ZIP: 01915

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/960,756

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Williams, Gregory D

; REGISTRATION NUMBER: 30901

; REFERENCE/DOCKET NUMBER: NEB-128

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 978-927-5054

; TELEFAX: 978-927-1705

; TELEX:

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1242 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: Genomic DNA

; FEATURE:

; NAME/KEY: Coding Sequence

; LOCATION: 1..1239

; OTHER INFORMATION:

; US-08-960-756-1

5821 GGGGCTTGGTGTAGAGGACCAGAAGACCC 5849

1849 TCAGGGGTACGAGAGTCTTGAACCTGCAAAAGTCTGAGGCTCTTCCCTCTCTCAG 1908
1002 TCAGGGGTACGAGAGTCTTGAACCTGCAAAAGTCTGAGGCTCTTCCCTCTCTCAG 943
1909 GCGATCCCGAGCGTAACCTTCTTCCGCGGACCCCGTCTTTGACGACCAATAGAGCC 1968
942 GCGATCCCGAGCGTAACCTTCTTCCGCGGACCCCGTCTTTGACGACCAATAGAGCC 883
1969 TTGAGCGTCTAGCTGCTGAGCTTCTCCGCGGATAGCGCAATCCGCTCCAGAGGGG 2028
882 TTGAGCGTCTAGCTGCTGAGCTTCTCCGCGGATAGCGCAATCCGCTCCAGAGGGG 823
2029 AAGTATCTCTCCGCAAGGCTTCCGCTAGGAGCCATCTTGGTTCTCCAGAGATGAC 2088
822 AAGTATCTCTCCGCAAGGCTTCCGCTAGGAGCCATCTTGGTTCTCCAGAGATGAC 763
2089 GGGATGGTGGTGTACGCTTCCCGCTTCTGCTGACAAAGGGGAAAGCTAGGATCTC 2148
762 GGGATGGTGGTGTACGCTTCCCGCTTCTGCTGACAAAGGGGAAAGCTAGGATCTC 703
2149 CTCTTCCGAAATAGGGGTAGCGGATTCGTTCCAAACGCTAGTCCCGCTTTTGGAGTAGAC 2208
702 CTCTTCCGAAATAGGGGTAGCGGATTCGTTCCAAACGCTAGTCCCGCTTTTGGAGTAGAC 643
2209 GAGGATCATGCTCTTTTCCGATCCGAGGCTTACGGGAAAGTTTGGATTTGAGC 2268
642 GAGGATCATGCTCTTTTCCGATCCGAGGCTTACGGGAAAGTTTGGATTTGAGC 583
2269 GATCGGCGGATATGTTTACGAAAGTTTCCGCGGCAAGACCTCATCAAGGATGAGCTT 2328
582 GATCGGCGGATATGTTTACGAAAGTTTCCGCGGCAAGACCTCATCAAGGATGAGCTT 523
2329 CACCTCGAACCCGATTTCTGCTATGTGAACGAAGATCAGTCTGAGTCCGCGCATCAG 2388
522 CACCTCGAACCCGATTTCTGCTATGTGAACGAAGATCAGTCTGAGTCCGCGCATCAG 463
2389 CTCCCTGAGAGTATCAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGT 2448
462 CTCCCTGAGAGTATCAGCGCTCCCTCAGGAACTCCACAAACTGAGGACCATCGAGGT 403
2449 GTCATCTGAGCGCCAACTGACCGTTTGGGTGCTGACGGTAGCAACGGGATCTGTTTC 2508
402 GTCATCTGAGCGCCAACTGACCGTTTGGGTGCTGACGGTAGCAACGGGATCTGTTTC 343
2509 ATCGCGCGCAAGAGAACTGCTGCGGTTTCCATAGGCGGGTCAATATAGACCACTG 2568
342 ATCGCGCGCAAGAGAACTGCTGCGGTTTCCATAGGCGGGTCAATATAGACCACTG 283
2569 GACCTTCCCGCATACCCACGAGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTC 2628
282 GACCTTCCCGCATACCCACGAGCTCCCGGAGCATCCACCGGAGAACCTGACCGTTTTC 223
2629 CCCCAAAAGTAGTGCAATAGGATCAATCTCAAAAGGGGGGCAATTCCTCCCTAGGAA 2688
222 CCCCAAAAGTAGTGCAATAGGATCAATCTCAAAAGGGGGGCAATTCCTCCCTAGGAA 163
2689 GAGGAGGTTCTTTTCCGCAAAAGTGTGGGTGGGCTGATCAAGATCTCTCTCTC 2748
162 GAGGAGGTTCTTTTCCGCAAAAGTGTGGGTGGGCTGATCAAGATCTCTCTCTC 103
2749 ATCGCGTTTCCGGGTAGACCAACCTAAAGGGGCAAGGTTCCGAGGTTTTCGAGGTTT 2808
102 ATCGCGTTTCCGGGTAGACCAACCTAAAGGGGCAAGGTTCCGAGGTTTTCGAGGTTT 43
2809 CAAGGGGCTTTTCGGGTCAACCGGCTAGCTAGGCTCAT 2850
42 CAAGGGGCTTTTCGGGTCAACCGGCTAGCTAGGCTCAT 1

RESULT 3

US-09-134-246-4/c

; Sequence 4, Application US/09134246B

; Patent No. 6207377

GENERAL INFORMATION:

APPLICANT: Wayne, Jay
APPLICANT: Xu, Shuang-yong
TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
TITLE OF INVENTION: Vectors And Identification Of Two Thermus Plasmid
TITLE OF INVENTION: Replication Origins
FILE REFERENCE: Thermus Shuttle Vector
CURRENT APPLICATION NUMBER: US/09/134,246B
CURRENT FILING DATE: 1998-08-14
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 1026
TYPE: DNA
ORGANISM: Thermus sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1023)
US-09-134-246-4

Query Match 17.5%; Score 1026; DB 3; Length 1026;
Best Local Similarity 100.0%; Pred. No. 9e-280;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3132 TCAGGAGGCTCTTGAAGGCTCTTCACTAGGTACGCTCCCGTCTGTCAGATGCCATC 3191
DB 1026 TCAGGAGGCTCTTGAAGGCTCTTCACTAGGTACGCTCCCGTCTGTCAGATGCCATC 967
QY 3192 CQTGTAATCTCGATAAACCCTGATAGGACGCAATAGATATTCGCGGGGAGTTCAAC 3251
DB 966 CQTGTAATCTCGATAAACCCTGATAGGACGCAATAGATATTCGCGGGGAGTTCAAC 907
QY 3252 CTTGGCCACAGCCCAAGCAAGCTTGCATAGAAAAGTCTTGAACGAGTCTTAGGAG 3311
DB 906 CTTGGCCACAGCCCAAGCAAGCTTGCATAGAAAAGTCTTGAACGAGTCTTAGGAG 847
QY 3312 ATCGGCAATGTAGTACGATGAGGCTGATAGGCGGAGTTCGAAACGCTCCACCTC 3371
DB 846 ATCGGCAATGTAGTACGATGAGGCTGATAGGCGGAGTTCGAAACGCTCCACCTC 787
QY 3372 GGGAGGACCAAGGATGAGCCCAAGCTTCAAGCCCAAGCTTTCGTTGGGATCAACCT 3431
DB 786 GGGAGGACCAAGGATGAGCCCAAGCTTCAAGCCCAAGCTTTCGTTGGGATCAACCT 727
QY 3432 TTTCCCTCAGCCCAAGGACCAAGCTTCAAGCCCAAGCTTTCGTTGGGATCAACCT 3491
DB 726 TTTCCCTCAGCCCAAGGACCAAGCTTCAAGCCCAAGCTTTCGTTGGGATCAACCT 667
QY 3492 GGCCTTGACCCAGTTGAGGAGAGCAGCCGCTTGGCCATGTCTAGGGCGAGTTCTCCA 3551
DB 666 GGCCTTGACCCAGTTGAGGAGAGCAGCCGCTTGGCCATGTCTAGGGCGAGTTCTCCA 607
QY 3552 GGGGTAGATGTAGTCTCAGGCTGAGCTTCCCTGCGCTCAGCGGACGCCCA 3611
DB 606 GGGGTAGATGTAGTCTCAGGCTGAGCTTCCCTGCGCTCAGCGGACGCCCA 547
QY 3612 AAGGTTGCGGATGCGCCCGGCTTCCCGTTGACGTTTGGTGAAGACGCTCGGTGGCCAC 3671
DB 546 AAGGTTGCGGATGCGCCCGGCTTCCCGTTGACGTTTGGTGAAGACGCTCGGTGGCCAC 487
QY 3672 CAGGCTTTTCTCAAGGACCTTCTTCCAGGCTGAGCGGCTGCGGGGTGACCCCCAG 3731
DB 486 CAGGCTTTTCTCAAGGACCTTCTTCCAGGCTGAGCGGCTGCGGGGTGACCCCCAG 427
QY 3732 GTTGGGGCCACATCTCCAGGGGACCATGAAGACGCTCCACCTTCTCCAGAG 3791
DB 426 GTTGGGGCCACATCTCCAGGGGACCATGAAGACGCTCCACCTTCTCCAGAG 367
QY 3792 CTCGCGTTCCGCTAGGGGATGTTGGAGCGGCAATCTCTCGAGAGTTCAGAGCTT 3851
DB 366 CTCGCGTTCCGCTAGGGGATGTTGGAGCGGCAATCTCTCGAGAGTTCAGAGCTT 307
QY 3852 CTGAGCGCGTCTCTGAGCGCTTGACATACATTCACCAAGGGGGGTTCAGCTGGTAT 3911

Db 306 CTGAGCCGCTCTGACGGCTTGACATACATTCCTCAACGGGGGTTGAGTGGTGTAT 247
QY 3912 AGGGGGGGCCCAACAGAGAAAGGAAGCCTCATGGTCTTTCTCGGTAAAGCAATCCT 3971
Db 246 AGGGGGGGCCCAACAGAGAAAGGAAGCCTCATGGTCTTTCTCGGTAAAGCAATCCT 187
QY 3972 AAGTGCTCTTTTGTGTATGTAAGCCCTTCGCGAGGCGATTTTCGGCACCTCCATCTGG 4031
Db 186 AAGTGCTCTTTTGTGTATGTAAGCCCTTCGCGAGGCGATTTTCGGCACCTCCATCTGG 127
QY 4032 AGGGGGGTCGGTGGCCCAAGAGTCTCTGACCCCTTATCTGACCCCTTATCTGAGTGGCATC 4091
Db 126 AGGGGGGTCGGTGGCCCAAGAGTCTCTGACCCCTTATCTGACCCCTTATCTGAGTGGCATC 67
QY 4092 GGTGTTGTCGTGGTGTCTCTAAAGCCCTCGTAAAGCTCTTCAAGAGAGGTTTTTTCGT 4151
Db 66 GGTGTTGTCGTGGTGTCTCTAAAGCCCTCGTAAAGCTCTTCAAGAGAGGTTTTTTCGT 7
QY 4152 CTTAC 4157
Db 6 CTTAC 1

RESULT 4
US-08-960-756-3
; Sequence 3, Application US/08960756
; Patent No. 5866422
; GENERAL INFORMATION:
; APPLICANT: WAYNE, JAY
; TITLE OF INVENTION: METHOD FOR CLONING AND
; TITLE OF INVENTION: PRODUCING THE Tsp45I RESTRICTION ENDONUCLEASE IN E. COLI
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: New England Biolabs, Inc.
; STREET: 32 Tozer Road
; CITY: Beverly
; STATE: MA
; COUNTRY: US
; ZIP: 01915
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,756
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Gregory D
; REGISTRATION NUMBER: 30901
; REFERENCE/DOCKET NUMBER: NEB-128
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 978-927-5054
; TELEFAX: 978-927-1705
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 999 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1...996
; OTHER INFORMATION:
; US-08-960-756-3

Query Match 17.1%; Score 999; DB 2; Length 999;
Best Local Similarity 100.0%; Pred. No. 3.8e-272;
Matches 999; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 614 ATGCAACAGATGCCCGAGTGGAAACGTGTGACACACAGAAAGCGTTGAGCTTCTGGAGAG 673
Db 1 ATGCAACAGATGCCCGAGTGGAAACGTGTGACACACAGAAAGCGTTGAGCTTCTGGAGAG 60
QY 674 GGGTATTTGGATAAACTACTGACAGTCTATAAAGGGGAAAGTGGCTCTTCGAGGTGAGTA 733
Db 61 GGGTATTTGGATAAACTACTGACAGTCTATAAAGGGGAAAGTGGCTCTTCGAGGTGAGTA 120
QY 734 CCAGAGAGGTGAGGAAACCTTCGCGAGGCTTCAAGGCATACGAGGGGAGGCGAGAT 793
Db 121 CCAGAGAGGTGAGGAAACCTTCGCGAGGCTTCAAGGCATACGAGGGGAGGCGAGAT 180
QY 794 AGTCCGGAGCGACAAACGAAACTCGTGGAGCGCTCTAAATGCCAGAAAGGTCGAG 853
Db 181 AGTCCGGAGCGACAAACGAAACTCGTGGAGCGCTCTAAATGCCAGAAAGGTCGAG 240
QY 854 CGGTCCCTTCAATCAACCCCTACCTGCTTGGTCTACTACTGCTTTCGAAAAAGCA 913
Db 241 CGGTCCCTTCAATCAACCCCTACCTGCTTGGTCTACTACTGCTTTCGAAAAAGCA 300
QY 914 GAAAGCGAACAAGGCCCTTGAGGAGGCAATTGACAGGCTTGCCTCAAGCACCAGAA 973
Db 301 GAAAGCGAACAAGGCCCTTGAGGAGGCAATTGACAGGCTTGCCTCAAGCACCAGAA 360
QY 974 ACCATCCGCTCTGCGCAAGGAGGCGCAAGAGAGGCGTAGAAGCTTGTATCCAAAG 1033
Db 361 ACCATCCGCTCTGCGCAAGGAGGCGCAAGAGAGGCGTAGAAGCTTGTATCCAAAG 420
QY 1034 CTCAGAGAGCTCCCGAATAAATCGGAGATAGGCGCGATGTTCAAAAGGTGTACAAA 1093
Db 421 CTCAGAGAGCTCCCGAATAAATCGGAGATAGGCGCGATGTTCAAAAGGTGTACAAA 480
QY 1094 GAAGAGCTAAAGGGGAAATAGAGAGAGGCTTCCAGGCGCTTACCAAAACCAAGATTGTG 1153
Db 481 GAAGAGCTAAAGGGGAAATAGAGAGAGGCTTCCAGGCGCTTACCAAAACCAAGATTGTG 540
QY 1154 GTAGTATCCCTGAAAAAGTAAACCGGAGCAAGCACCCCTTATTCGCGAGAGAGAGCG 1213
Db 541 GTAGTATCCCTGAAAAAGTAAACCGGAGCAAGCACCCCTTATTCGCGAGAGAGAGCG 600
QY 1214 GGCATCATATATACACGGATCGGATGAGCTTTCARAGATGCCGCCAAGGAAACCTTG 1273
Db 601 GGCATCATATATACACGGATCGGATGAGCTTTCARAGATGCCGCCAAGGAAACCTTG 560
QY 1274 GGCCTTGGCGAGGAAGCAGAACTAGGACCAAGGGCGTAGATTCTACGTGTCATCCGG 1333
Db 661 GGCCTTGGCGAGGAAGCAGAACTAGGACCAAGGGCGTAGATTCTACGTGTCATCCGG 720
QY 1334 CGTAGCCTGAAGACATGCGACCTTAAACAGAGAGTGAAGTTCAATCCGACTTGGC 1393
Db 721 CGTAGCCTGAAGACATGCGACCTTAAACAGAGAGTGAAGTTCAATCCGACTTGGC 780
QY 1394 GGAACCAAGACAAACAGAACTAGTAGCAAGGCTTCCATAGGTTGGACCTTGAGAAG 1453
Db 781 GGAACCAAGACAAACAGAACTAGTAGCAAGGCTTCCATAGGTTGGACCTTGAGAAG 840
QY 1454 AGGCATAGGAATAGTGTGTTGAGCGGAATGCTGTGTGAGGCAAGTTTGTGGGTGG 1513
Db 841 AGGCATAGGAATAGTGTGTTGAGCGGAATGCTGTGTGAGGCAAGTTTGTGGGTGG 900
QY 1514 GCCGGACTGGGAAAGAAACGATCGTTACATCGTACTCTCTTCCAGACCTGTATAGG 1573
Db 901 GCCGGACTGGGAAAGAAACGATCGTTACATCGTACTCTCTTCCAGACCTGTATAGG 960
QY 1574 GAGCTCTACCAAGGGTGAAGAGCCCTGGGCTCTAG 1612
Db 961 GAGCTCTACCAAGGGTGAAGAGCCCTGGGCTCTAG 999

```
RESULT 5
US-09-134-246-5/c
; Sequence 5, Application US/09134246B
; Patent No. 6207377
; GENERAL INFORMATION:
; APPLICANT: Wayne, Jay
; APPLICANT: Xu, Shuang-yong
; TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
; TITLE OF INVENTION: Vectors And Identification Of Two Thermus Plasmid
; TITLE OF INVENTION: Replication Origins
; FILE REFERENCE: Thermus Shuttle Vector
; CURRENT APPLICATION NUMBER: US/09/134,246B
; CURRENT FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 180
; TYPE: DNA
; ORGANISM: Thermus sp.
US-09-134-246-5

Query Match          3.1%; Score 180; DB 3; Length 180;
Best Local Similarity 100.0%; Pred. No. 5.5e-41;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4128 CTCCTCAAAGAGGTTTTCGTTTCACCTCGGACCTCCTGTGTCATCTGGAGCCGA 4187
DB 180 CTCCTCAAAGAGGTTTTCGTTTCACCTCGGACCTCCTGTGTCATCTGGAGCCGA 121
QY 4188 GCGCTTACCTAGCTCTGGGGGTGATCCGGGGCAACGCGTTCGCTTTTATG 4247
DB 120 GCGCTTACCTAGCTCTGGGGGTGATCCGGGGCAACGCGTTCGCTTTTATG 61
QY 4248 GGTCCAAATAACCGTCAGCCAGCGGCTGGCAATCCCCCTCTCTAAAGCCGTTATG 4307
DB 60 GGTCCAAATAACCGTCAGCCAGCGGCTGGCAATCCCCCTCTCTAAAGCCGTTATG 1

RESULT 6
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
```

```
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pt-Fls
US-08-232-463-14

Query Match          1.3%; Score 77; DB 1; Length 7218;
Best Local Similarity 4.9%; Pred. No. 7.5e-11;
Matches 20; Conservative 240; Mismatches 145; Indels 0; Gaps 0;

QY 469 GATTCAGTATACCAATAATGCGACAAAGTTTGTGAGAGGTGTCACAAACAAAGGCTTT 528
DB 1445 GAATTTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1386
QY 529 CTCGGTCAGTTATGTTGAGGTGGGGCGGTCAAAAGGCGCACTTAAGTTTGGTAAAGCG 588
DB 1385 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1326
QY 589 GGAGGAAGCAACCGGGGTGTTACATGCAACAGATGCGCGAGTGGAACGTGTGGACACA 648
DB 1325 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1266
QY 649 GAGACGCTGAGCTTCTCGAGAGGGGTATTGGATAAATCTGTCAGGTCTATAAAGG 708
DB 1265 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1206
QY 709 GGAAAGTGCTCTTCGAGGTACGTACGAGAGAGGTAGAGAAACACTTCGCGAGGCGCTA 768
DB 1205 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1146
QY 769 CAAGGCATAGGGGAGGAGGAGGATAGTCGGAGGAGCAACGAACTCGTGAAGCCGT 828
DB 1145 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1086
QY 829 GCTAAATGCCAGAAAAGGTGCGAGCGTCCCGCTTCAATCACCC 873
DB 1085 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1041

RESULT 7
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
```


QY 3560 TGATGTCGTCAGGTCAGCTGCTTTCCCTGCGCTCACCAGCGGCCCAAGGGTGC 3619
Db 797 CCTCTTCGGTCCGACAGCGGATCGCTCCCGCGGTATACATCCCGCCAGCGCGAGTGT 738
QY 3620 CGATGGCCCGCGCTCCCGTTGACGCTTTGGTCAAGGAGCTCGGTGGCCACAGGCGCT 3679
Db 737 CGATCGCAGCTGTGGCTTCTCGCTGCATCGCGCGCGCGGGGTGGCTTCGAGGCTGT 678
QY 3680 TTTTCTCAAGACCTTCTCCAGCGTGGAGCGGTCTGCCGGGTGACCCCGAGTTGAGGG 3739
Db 677 CGGCTCGACGCTGCTTCCGCGCGGTGACGATGTTGAGGTTCTTCGCGTAGAGCTTGGCGT 618
QY 3740 CCAACATCTCCAGGGGACCATGAGAGCAGCGTCCCGCTTCCAGAGCTCCCGGT 3799
Db 617 TGAGCTTGGCGCTCGGCTGATCAGGTGCAACTGTTCCGAGGTGCGAGCTTCAGTTCGG 558
QY 3800 TGGCGTAGGGGATGCTGAGCGGCGCAATCTCTGGAGGAGTTCAGAGAGCT 3850
Db 557 CGCCTTCGACGACGATGTCGCGCGCTCCACCTGGAAGCGCTCCAGGCGCT 507

RESULT 10

US-09-252-991A-15763

; Sequence 15763, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 15763

; LENGTH: 1953

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-15763

Query Match 0.9%; Score 51.8; DB 4; Length 1953;

Best Local Similarity 44.4%; Pred. No. 0.0004;

Matches 209; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

QY 3380 CCAGGATGAGCCCGAGTCAACGCGCCAGCTTGGTGTGGGCGATCACCTTTCCCT 3439

Db 942 CCGCGCTCGGATCTCGCGCTGCGCGGTGCGCGGTAGGTCTTCCGCTTCAGCTCCACGCGCT 1001

QY 3440 GAGCCAGAGACACGACGCTCCAGGTTGGGGCGGATTCGTTGCTCTGTAGGCGCTTGA 3499

Db 1002 GGGCGCGATTTTCTAGCTGCGCTTGGTGGAGGCTGGGCGAGCTCAGCTTCCCGTGG 1061

QY 3500 CCCAGTTGAAGAGAGACGCGTTGGCCATGTTAGGGGAGGTTCTTCAAGGGTGA 3559

Db 1062 CGTCAGTCGATGTGCGCGCGCTGCGCGGCATGTGCGCGGCAGCGCACCCCGCACGC 1121

QY 3560 TGATGCTGTCAGGTGAGCTGCTTTCCTGCGCTCAGCGGACGCGCCCAAGGGTGC 3619

Db 1122 CTTGTTCCGTCGCCACAGGCGGATTCGCCCCGGGTATATCCCGCGCGTTCAGCTTCC 1181

QY 3620 CGATGGCCCGCGCTCCCGTTGACGTTTGGTGAAGGAGCTGCGGTGGCCACAGGCGCT 3679

Db 1182 CGATCGCCAGCTGTGGCTTCTCGCTGCATCGCGCGCGCGGGGTGGCTTCGAGGCTGT 1241

QY 3680 TTTTCTCAAGACCTTCTTCCAGCGGTGGAGCGTCTCCCGGTGACCCCGAGTTGAGG 3739

Db 1242 CGGCTTGACGCTGCTTGGCGCGGTGACGATGTTGAGTCTTTCGCGTAGAGCTTGGCT 1301

QY 3740 CCAACATCTCCAGGGGACCATGAGAGCGCGTCCCGCTTCCAGAGCTCCCGGT 3799

Db 1302 TGAGCTTGGCGCTCGGGTGATCAGTCAACTGTTGAGGTTGCCGAGTTCAGTTCGG 1361
QY 3800 TCCCTAGGGGATGTCGAGCGGCAATCTCTGAGGAGTTCAGAGCT 3850
Db 1362 CGCCTTCGACGACGATGTCGCGCGCTCCACTGGAAGCGTCCAGGCGCT 1412

RESULT 11

US-09-252-991A-15871/c

; Sequence 15871, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 15871

; LENGTH: 3351

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-15871

Query Match 0.9%; Score 51.8; DB 4; Length 3351;

Best Local Similarity 44.4%; Pred. No. 0.0062;

Matches 209; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

QY 3380 CCAGGATGAGCCCGAGTCAACGCGCCAGCTTGGTGTGGGCGATCACCTTTTCCCT 3439

Db 1417 CCGCGCTCGGATCTCGCGCTGCCACGCGCTAGTCTTTCGCTTACCTCCAGGCT 1358

QY 3440 GAGCCAGAGACACGACGCTCCAGGTTGGGGCGGATTCGTTGCTCTGTAGGCGCTTGA 3499

Db 1357 GGGCGCGATCTTCAGGTCGCCCTGGTGGAGGCTGGGCCAGGCTCAGCTTGGCGCTGG 1298

QY 3500 CCAGTTGAAGAGAGACGCGCTTGGCGATCTTAGGGGAGGTTCTTCCAGGGTGA 3559

Db 1297 CGTCAGTCGAGTGTGCGCGCGCTGGCGCGCATGTCGCCCGCGCACCCCGACGC 1238

QY 3560 TGATGTCGTCAGGTGAGCTTTCCTTGGCTCAGCGGACGCGGCCCAAGGGTGC 3619

Db 1237 CTTGTTCCGTCGCGAGCGGATGCGCCCGCTACATCCCGCGCGAGCTGT 1178

QY 3620 CGATGGCCCGCGCTCCCGTTGACGTTTGGTGAAGGAGCTCGGTGGCCACAGGCGCT 3679

Db 1177 CGATCGCCAGCTGTGGCTTCTCGCTGCCATCGCGCGCGCGGCTGGCGCTTCAGGCTGT 1118

QY 3680 TTTTCTCAAGGACCTTCTTCCAGGGTGAGAGCGTCTCGCGGTGACCCCGAGTTGAGG 3739

Db 1117 CGGCTTGAGCTGTTCGCGCGGTGACGATGTTGAGGTTCTTCGGTAGAGCTTGGCT 1058

QY 3740 CCAACATCTCCAGGGGACCATGAAGAGACCGTCCCGACCTTCTTCCAGAGCTCCCGGT 3799

Db 1057 TGAGCTTGGCGCTCGGGTGATCAGTGCAGTGTTCGAGGTTGCCGAGCTTCAGTTCGG 998

QY 3800 TGGCGTAGGGGATGTCGAGCGGCAATCTCTGAGGAGTTCAGAGCT 3850

Db 997 CGCCTTCGACGACGATGTCGCGCGCTCCACTGGAAGCGCTCCAGGCGCT 947

RESULT 12

US-09-252-991A-5348/c

; Sequence 5348, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS


```
Db 2780267 ACATTTCCCGCTTAGTGGTGGTCTCGTCGAGCAACACAGGTGCGGTGTTGGAC 2780326
Qy 3424 ATCAACCTTTTCCCTGAGCCAGAGACACGACGCTCCAGGTGGGGCGGATTCGGTGG 3483
Db 2780327 AACAGCAGTTTGACAGCGCCACCCGGCGACGCTCGCCACCGGATAGTGTGGTTACCGG- 2780385
Qy 3484 TCCTGGTAGCGCTTGACCCAGTTGAAGGAGAGACGCGCTGGCCATGTCTAGGGCGAGG 3543
Db 2780386 --CTCGTCGGCGCGGACAGCGACGCGATCCATGAGCTGTCTGAGCTCGCGGTGAGG 2780443
Qy 3544 TTCTCCAGGGTAGATGTAGTCTCTCCAGGTGAGCTGTTCCTTCCCTGCGCTCAGCCGG 3603
Db 2780444 TCCACGCGTCCGGGTGCTCCAGTCTCTTTCGACCGACCCATCTCTCCATCAGCTCG 2780503
Qy 3604 ACGCCCAAGAGGTGCGATGGCCCGCGCTCCCGCTTACGCTTGGTGAAGSACGTG 3663
Db 2780504 TCGGTGTAGTCGGTGGCCATCAATTCGGCGACCTTCGTTGAAGCGGTGAGCTTGATCTTG 2780563
Qy 3664 GTGGCCACCAAGGCCCTTTTCTCAAGGACCTTCTTCCAGCGGTGGACGCTCTCCCGGGTG 3723
Db 2780564 ATGTCCCCCATGCGCTCTTCCACATTCGCGAACGGTCTTGCTCGTTCAGCGCGGT 2780623
Qy 3724 ACCCCAGGTGAGGGGCAACATCTCCAGGGGACCATGAAGAC 3767
Db 2780624 TCCTGTTGAGGATGCCCAACGGTGGCGCGGTGGCCAGGAAGGC 2780667
```

RESULT 15

```
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37RV
US-09-103-840A-1
```

```
Query Match 0.8%; Score 49.6; DB 3; Length 4411529;
Best Local Similarity 45.7%; Pred.No.0.23;
Matches 212; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

Qy 3304 TCTAGGAGATCGCAATGCTAGTACCAATGAGGTGATAGGCGCGGAGTTTGGAAACGC 3363
Db 2783397 TCGTGGGTGACCGCAAGATCGACCGGGTAGCTGCCAGATCTCTTTCGAGCCACTGC 2783456
Qy 3364 TCCACCTCGGGGAGGACAGGATGAGGCCAGGTCAACGGCCACGGTCTTGGTGTGGGC 3423
Db 2783457 ACATTTCCCGCTTAGTGTGGTTCGCTCGTGGCAACACAGGTGCGGTTCGAC 2783516
Qy 3424 ATCAACCTTTTCCCTGAGCCAGGACGACGCTCCAGGTGGGGCGGATTCGGTGG 3483
Db 2783517 AACAGCAGTTTGCAACAGCGCCACCGCGCGACGCTCGCCACCGGATAGGTGGTTACCGG- 2783575
Qy 3484 TCCTGGTAGGCTTGACCCAGTTGAAGGAGAGACGCGCTTGGCCATGTCTAGGGCGAGG 3543
Db 2783576 --CTCGTCGGCGCGGACGACGCGATCCATGCGCTCTCGAGTGGCGTCCGAGG 2783633
Qy 3544 TTCTCCAGGGTAGATGTAGTCTCGTCCAGGTGAGCTGCTTCCCTGCGCTCAGCCGG 3603
```

```
Db 2783634 TCCACAGGTGCGCGTGGTCCAGTTCCTTTTCAGCCGACCCATCTCTTCCATCAGCTCG 2783693
Qy 3604 ACGGCCCCAAGGTGCGATGCGCCCGCGCTCCCGCTTTCAGCGTTTGGTGAAGGACGTG 3663
Db 2783694 TCGGTGTAGTCCGTGGCCATCAATTCGGCGACCTCGTTGAAGCGGTGAGCTTGATCTTG 2783753
Qy 3664 GTGGCCACCAAGGCCCTTTTCTCAAGGACCTTCTTCCAGGGGTGGACGCTCTGCGGGTG 3723
Db 2783754 ATGTCCCCCATGCGCTCTTCCACATTCGCGAACGGTCTTGTCTCTGTTTCAGCGCGGT 2783813
Qy 3724 ACCCCAGGTGAGGGGCAACATCTCCAGGGGACCATGAAGAC 3767
Db 2783814 TCCTGTTGAGGATGCCCAACGGTGGCGCGGTGGCCAGGAAGGC 2783857
```

Search completed: March 18, 2004, 05:51:35
Job time : 287.095 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 02:39:53 ; Search time 1345.06 Seconds

(without alignments)
16085.819 Million cell updates/sec

Title: US-09-664-186-6

Perfect score: 5849

Sequence: 1 tctagaagtcagggtggac.....ttagagaccagaagacc 5849

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications NA:

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length DB	ID	Description
1	53.8	0.9	9025608	14	US-10-156-761-1
C 2	53.2	0.9	1755	15	US-10-260-238-721
C 3	51.6	0.9	1815	9	US-09-758-269-13
C 4	50.8	0.9	1018	12	US-10-425-114-35941
C 5	50.8	0.9	1077	14	US-10-156-761-6573
C 6	50.8	0.9	9025608	14	US-10-156-761-1
C 7	50	0.9	675	14	US-10-156-761-6076
C 8	49.8	0.9	1389	14	US-10-156-761-5431
C 9	49.6	0.8	1674	12	US-10-282-122A-26289
C 10	49.6	0.8	1677	12	US-10-282-122A-28540
C 11	49.4	0.8	898	12	US-10-425-114-32815
C 12	49.2	0.8	1803	15	US-10-369-493-42225
C 13	48.8	0.8	1045	12	US-10-425-114-26418
C 14	48.6	0.8	9369	15	US-10-369-493-34210
C 15	48.6	0.8	9369	14	US-10-200-562-190

```
C 16 48.6 0.8 9369 14 US-10-237-551-190
C 17 48.6 0.8 9369 14 US-10-237-551-247
C 18 48.6 0.8 154746 10 US-09-827-688-8
C 19 1977 14 US-10-355-430-17
C 20 47.8 0.8 963 12 US-10-282-122A-11900
C 21 47.6 0.8 669 12 US-10-425-114-34873
C 22 47.6 0.8 864 12 US-10-425-114-35445
C 23 47.6 0.8 1527 14 US-10-156-761-6366
C 24 47.4 0.8 1326 14 US-10-156-761-2557
C 25 47.2 0.8 969 14 US-10-156-761-4891
C 26 46.8 0.8 1224 14 US-10-156-761-3605
C 27 46.8 0.8 1605 15 US-10-260-238-908
C 28 46.8 0.8 2223 14 US-10-156-761-3605
C 29 46.6 0.8 1071 15 US-10-369-493-42984
C 30 46.6 0.8 1065 14 US-10-160-758-1
C 31 46.6 0.8 1065 15 US-10-085-117-114
C 32 46.6 0.8 10455 14 US-10-160-758-4
C 33 46.6 0.8 11073 14 US-10-160-758-2
C 34 46.6 0.8 11073 15 US-10-085-117-113
C 35 46.4 0.8 1381 12 US-10-425-114-13514
C 36 46.2 0.8 1414 12 US-10-425-114-14100
C 37 46.2 0.8 1473 9 US-09-853-161-43
C 38 46.2 0.8 1473 9 US-09-852-659A-43
C 39 46.2 0.8 1473 9 US-09-852-797-43
C 40 46.2 0.8 2892 14 US-10-156-761-2448
C 41 46 0.8 808 12 US-10-424-599-52912
C 42 46 0.8 1263 15 US-10-369-493-40502
C 43 46 0.8 1418 12 US-10-425-114-1960
C 44 46 0.8 1532 12 US-10-425-114-34918
C 45 46 0.8 1864 12 US-10-425-114-28403
```

ALIGNMENTS

RESULT 1

```
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication NO. US20030119018A1
; GENERAL INFORMATION:
```

```
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
```

```
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
```

Query Match 0.9%; Score 53.8; DB 14; Length 9025608;
Best Local Similarity 45.0%; Pred. No. 0.01;
Matches 202; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

```
QY 3449 GGACGACGACGTCAGGGTGGGCGGATTCGGTGTCTGTGTAGCCTTGACCCAGTGA 3508
DB 4922065 GGTGTGCGCGCGACCCAGAGCGGTCATCGTGGCCCGATCTCTGTCGACCCAGATGA 4922124
```

Matches 127; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 3366 CACCTCGGGAGACACAGATGAGGCCAGGTCAACGGCCAGGCTCTTGGTCTGGGAT 3425

Db 495 CACCTCGGGAGACACAGATGAGGCCAGGTCTTGGTCTGGGAT 436

QY 3426 CACCTCTTTCCCTGAGCCAGAGGACACAGCTCCAGGCTGGGGCGGATTCGGTGTTC 3485

Db 435 CTCTGCTTACATGACCTCCAGCGTGGCGGGCTCGACCGAACCTGAATTTGTTC 376

QY 3486 CTGCTAGGCTTACCCAGTGAAGAGAGACCGCGTTGGCCATCTTAGGGCGAGTT 3545

Db 375 GTGGACGGTCTGACCGTATGAGGGGAGGTGGAGATCGGATGACGAGGTGGAGAC 316

QY 3546 CTTCCAGGGGTAGATGATGCTCCAGGCTGAGCTGGGCTTTCCCTGGGCTCAGCCGGAC 3605

Db 315 GTTCCGCTGGAGACCGCTAGTCCATGATGGGCTGTCTTCCGGGCGAGCTCCCGGC 256

QY 3606 GGCCCAAGG 3615

Db 255 ATCGAGACG 246

RESULT 3

US-09-758-269-13/c

; Sequence 13, Application US/09758269

; Patent No. US20020104120A1

; GENERAL INFORMATION:

; APPLICANT: IUCHI SATOSHI

; APPLICANT: KOHAYASHI, MASATOMO

; APPLICANT: SHINOZAKI, KAZUO

; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN

; TITLE OF INVENTION: CLEAVAGE ENZYME GENE

; FILE REFERENCE: 3914-3

; CURRENT APPLICATION NUMBER: US/09/758,269

; CURRENT FILING DATE: 2001-01-12

; PRIOR APPLICATION NUMBER: JP 2001-003476

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: JP 2000-010056

; PRIOR FILING DATE: 2000-01-13

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 13

; LENGTH: 1815

; TYPE: DNA

; ORGANISM: Zea mays

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1812)

US-09-758-269-13

Query Match 0.9%; Score 51.6; DB 9; Length 1815;

Best Local Similarity 45.9%; Pred. No. 0.00035;

Matches 177; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 3453 CAGCAGCTCCAGGGTGGGGCGGATTCGGTGGTCTCTGAGGCTTACCCAGTTGAAGGA 3512

Db 540 CAGCGCTGCACCATCCCTGCGCGTGAAGAGTGTGCCCCCGACGGGTTCGAGCA 481

QY 3513 GAGCAGCCCTTGGCCATGTCTAGGGCGAGGTTCTCCAGGGGTAGATGATGCTCCAG 3572

Db 480 GGGTTTGGCGCGTTGGCGCGTAGACCCCGTAGACCGGCGGATGAGGCGCGGAGAC 421

QY 3573 GGTGAGCTGGCTTTCCTGGCTCAGCGGACGGCCCAAGGGTGCAGATGCCCGCG 3632

Db 420 GGGGAGCTCGTGCACGGGCGGCTCTCCCGACGGGCGGAGTTGGCGCGATCTGCAC 361

QY 3633 CTCCTCCCTTTCAGCGTTTGGTGAAGAGAGCTCGGTGGCCACAGGCCCTTTTCTCAAGGAC 3692

Db 360 GCGCGGTTCGCGCTGCTGGGCGGCTCGAGGACGTTGGCGGCTTGGCCACGACCC 301

QY 3693 CTTCTCCAGGCTGAGCGTCTGCGGGGTGACCCCGGTTGAGGCGCCACATCTCCAG 3752

QY 3509 AGGAGACACGCGTTGGCCATGTCTAGGGCGAGGTTCTTCCAGGGTAGATGATGCT 3568

Db 4922125 ACGGCGGTGGCGGCCACTCGACAGATGCGGGTCTTCCAGTGCAGGCGCGCGGC 4922184

QY 3569 CAGGGTAGCCTGGCTTTCCTGCGCTCAGCGGACGGCCCAAGGGTGGCGATGGGCC 3628

Db 4922185 CTTGTTGGGTGAGGCTTCCGCGCACTCCACACAGGAGCGCTGGGCACTCCGATCCG 4922244

QY 3629 GCGGCTCCCGTTGACGGTTTGGTGAAGGACCTCGGTGGCCACCAAGGCGCTTTTCTCAA 3688

Db 4922245 GCGGACGACACTGCTTGGGCTATTATCCAGTTCGAGGCCACACAGCTCAGACCCCA 4922304

QY 3689 GACCTTCTTCAGGGGTGAGCGGTGCGCGGTGACCCCGAGGTTGAGGGCCACATCT 3748

Db 4922305 GTCCATCAGACGCGTCCGACGGCTGTGTCAGCGCGCCAGCAGCGGTTCCCTCGT 4922364

QY 3749 CCAGGGGACCATGAGACGACCTCCCACTCTTCCAGAGCTCCCGGTGCGGTAGG 3808

Db 4922365 CGAGAACCCACGAGCGCGGCGGACGCGGCTGAGCGCGGACGCGCCGACG 4922424

QY 3809 GGATGTTGAGCGGGCAATCTCTCTGAGAGGTTTCAGAGCTTCTGAGCGCGCTCTGGA 3868

Db 4922425 GGATGAGGGTCCGTGACGTGAGGAAAGACGCGGCGCTCCACAGAGCGGTCATGA 4922484

QY 3869 CGCTTTCATACATTCCTCCCAAGCGGGGT 3897

Db 4922485 CGCACGATAACCTCGCGGGGACGGGT 4922513

RESULT 2

US-10-260-238-721/c

; Sequence 721, Application US/10260238

; Publication No. US20040016025A1

; GENERAL INFORMATION:

; APPLICANT: Budworth, Paul R.

; APPLICANT: Noughamer, Todd G.

; APPLICANT: Briggs, Steven P.

; APPLICANT: Cooper, Bret

; APPLICANT: Glazebrook, Jane

; APPLICANT: Goff, Stephen A.

; APPLICANT: Katagiri, Fumiyaki

; APPLICANT: Kreps, Joel

; APPLICANT: Provart, Nicholas

; APPLICANT: Ricke, Darrell

; APPLICANT: Zhu, Tong

; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION

; FILE REFERENCE: 6011-NP

; CURRENT APPLICATION NUMBER: US/10/260,238

; CURRENT FILING DATE: 2002-09-26

; PRIOR APPLICATION NUMBER: US 60/325,448

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/325,277

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 60/370,620

; PRIOR FILING DATE: 2002-04-04

; NUMBER OF SEQ ID NOS: 6077

; SEQ ID NO 721

; LENGTH: 1755

; TYPE: DNA

; ORGANISM: Oryza sativa

; FEATURE:

; NAME/KEY: N region

; LOCATION: (917)..(917)

; OTHER INFORMATION: n = any nucleotide

; FEATURE:

; NAME/KEY: N region

; LOCATION: (1750)..(1750)

; OTHER INFORMATION: n = any nucleotide

US-10-260-238-721

Query Match 0.9%; Score 53.2; DB 15; Length 1755;

Best Local Similarity 50.8%; Pred. No. 0.00011;

```

Db      300  TTCTCTGAAACGGCTCGAGGGCGGCGCGCGCGCGCTGGAAACAAGTTGAGCTGCTCTT 241
QY      3753  GGGGACCATGAAGACGACCGTCCCAACCTTCTCCAGAGCTCCCGGTTGCCGTAGGGGAT 3812
Db      240  GCGCGCCTCCGCTTTCTCTGGCGCGCGCGCGCTGCCCTGGACGCAATGGCGCGCGGCTT 181
QY      3813  GGTGGAGCGGCAATCTCTCTGAGGA 3838
Db      180  CTGGACGCGCGCAGGTCGGCGA 155

RESULT 4
US-10-425-114-35941/c
; Sequence 35941, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Planes and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425.114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 35941
; LENGTH: 1018
; TYPE: DNA
; ORGANISM: Zea mays subsp. mexicana
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMROTEOSINTE020D09_FLI
US-10-425-114-35941

```

Query Match.	0.9%;	Score 50.8;	DB 12;	Length 1018;
Best Local Similarity	46.6%;	Pred. No. 0.00043;		
Matches 163;	Conservative 0;	Mismatches 187;	Indels 0;	Gaps 0;
Qy	3525	GGCATGCTAGGGCGAGGTTCCTCCAGGGGTAGATGATGCTGCCAGGTGAGCCTGGC	3584	
Db	738	GTCTTTGTGGTGTTGTGTGTAGTGTGTGGCTGCGCTGCTGTGTGTGCCCGCC	679	
Qy	3585	TTTCCCTGGGCTCAGCCGCGACGGCCCAAGAGGTGCCGATGCCCGCGCTCCCGTTGAC	3644	
Db	678	CATCCCGACGGTGCTCATCACCGGTCTTGGCGCGCCGCGCGCTCTCACCTGCTC	619	
Qy	3645	GGTTTGGTGAAGACGTCGTGTGGCCACAGCGCTCTTTTCTCAAGACCTTCTTTCAGGC	3704	
Db	618	AGTGGCCTGTGTGATGACGCGCGCGGACTGTCTCTTCGCGAGACGCGGTGTCTTTGGC	559	
Qy	3705	GTGACCGGTCTGCGGGTGACCCCGACGGTTGAGGGCCAACTCTCCAGGGGGACCATGAA	3764	
Db	558	GTACTGCCCGCCCTGCGCGGCTTCTGCTTGGCGGCTCTCATGGCGTGCCTTCTG	499	
Qy	3765	GACACCGTCCCACTTCCTCCAGAGCTCCCGTTCGCCGTAGGGGATGTTGGAGCGGGC	3824	
Db	498	CTTGGCGGCTCCGTGCTCTCGCGGCTTCTGCTTGGCGGCTCCGCTGCTCTCGCGGC	439	
Qy	3825	AATCTCTGGAGGAGTTCCTCAAGACTTCTGAGGCCCGTCTCTGGACGGCTT	3874	
Db	438	CTTCTGCTTGGCGGCTTCCTGCTGCTCGCGGCTTGTGTGTTGGCGGCTT	389	

RESULT 5
US-10-156-761-6573/c
; Sequence 6573, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN

```

RESULT 6
US-10-1566-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, NASHAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262

```

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 0.9%; Score 50.8; DB 14; Length 9025608;
Best Local Similarity 44.3%; Pred. No. 0.082;
Matches 208; Conservative 0; Mismatches 262; Indels 0; Gaps 0;
QY 3366 CACCTCGGGAGGACAGATGAGCCCAAGGTCAAGCCACGGCTTGGTGTGGGAT 3425
Db 7895885 CGCGCGGGGATGCGGTGAGGGCGATCCGCTCCGTACACGGCGGAGTCCAGCG 7895826
QY 3426 CACCTTTTCCCTGAGCCACAGGACACGACGTCAGGGTGGGGCGGATTCCTGTGTC 3485
Db 7895825 CCGCGCCGTCGTGTCGCGCAGGACCGGAGTCTTTCGCGGTCGAGTTCGCGTA 7895766
QY 3486 CTGGTAGGCTTACCCAGTTGAGGAGACGCGGTCGTCATGTAGGGCGAGTT 3545
Db 7895765 CACGACGCGGACAGGGTGGCGCCCGAGTGGAGATCTCAGCGGCTTGAACACACCTG 7895706
QY 3546 CTTCCAGGGGTAGATAGTGTGTCAGGGTGAAGTCTTCCCTGGGCTCAGCGGAC 3605
Db 7895705 CTGGTCTTGGCCGATGCCAGACGACGTTGTGGCGGCCACCGCGGTGATCCCA 7895646
QY 3606 GGCCCAAGGTCGATGCGCGGCTCCCGGTGAGCGTTTGGTAGAGACGTCGGT 3665
Db 7895645 GCGGTGCGGATGTCACGCGCGGCGCCACGACCTCAGCGGAGCTCCACCCCTGTT 7895586
QY 3666 GGCCACAGGCTTTTCTCAGGACCTTCTCAGGGGTGAGGTCGTCGGGTGAC 3725
Db 7895585 GTGGTAGCCCGCGATCTCGCGGCGGTGTCGAGGGGACGAGTAGTCGGTGGC 7895526
QY 3726 CCGCAGGTTGAGGGCAACATCTCCAGGGGGACCATGAAGACGACCGTCCCGCTTCT 3785
Db 7895525 CCGCGCCGCGGCGGAGCGACTCTTCTCGCGGAGAGCTCCACCGGACGATCTTGA 7895466
QY 3786 CAGAGCTCCCGTTCGTTAGGGATGAGGCGGGCAATCTCTGA 3835
Db 7895465 CGCGCCCGGATCCGGGCGCGCTGGAGGCTCGGAGGCGGACCCCGCGGA 7895416

RESULT 7
US-10-156-761-6076/c
; Sequence 6076, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-05-30

; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6076
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(675)
US-10-156-761-6076

Query Match 0.9%; Score 50; DB 14; Length 675;
Best Local Similarity 44.7%; Pred. No. 0.0006;
Matches 237; Conservative 0; Mismatches 290; Indels 3; Gaps 1;
QY 3241 GGGAGTTTCACTCCCTGGCCACAGCCCAAGCTGCAAGAAAGTCTTGAACGACGG 3300
Db 632 GGGTGTCTCCCTCGGTGATCGCGCGACGACACGCGGGCGGCTTCGAGA 573
QY 3301 TCATCTAGGAGATCGCAATGTACGTAGCAATAGGGTGTATAGGGCCGGAGTTGGAA 3360
Db 572 ACCTGTGAGATTCGCGAGTCCATGCCGCCGATGGCGAACCCAGGGGCGGTTCGTGCC 513
QY 3361 CGTCCACCTCGGGGAGGACGAGTGGGCCCGAGTCAAGCCACGCTTGTGTGTTG 3420
Db 512 AGGGCGGCTGTGGCGGACCGAGTCCAGGCCCGGGCGGTGCCGCCGGGCTTGTGGGG 453
QY 3421 GGCATCACCTTTTCCCTGAGCCACAGACACGACGTCAGGGTGGGGCGGATTCGG 3480
Db 452 GTGGCCAGCAGGACCCGTCAGAAAGTAGTCCACGCCCTCTGAGACGCGCGGCTCG 393
QY 3481 TGGTCTGTAGGCTTGACCCAGTTGAAGGAGACGCGCGTTCGCTGTCATGTCTAGGGG 3540
Db 392 GCCTCGGCTTCGAGTCCGTTGAGCGGCGGATGAGACGTCGCGCGCGGAGTTCGCCCG 333
QY 3541 AGGTTCCTCCAGGGTAGATGTAGTCCGTCAGGGTGAAGCTTCTCCCTGGCTTCAGC 3600
Db 332 CGCGCGGACGCGGAGTCCGCTCCCGAGTGGAGACGTCGAGTCCGAGTCCGCGGTGC 273
QY 3601 CGGACGCGCCAAAGGTCGATGGCGCGCTCCCGTTCGAGCGTTTGGTGAAGGACG 3660
Db 272 GCGAGTCCCGCGTTCGTTGACCGGAGAGTTCGCTGGCGG---CGACAGGCGTCC 216
QY 3661 TCGGTGCGCACCGGCGCTTTTCTCAAGGACCTTCTCCAGGCGTGGAGCGTCTCCCG 3720
Db 215 GCGAAGACCTGAGGTGCTCCAGCTCTCTCGCGCGCTTCCATGCTTGTGCGCACCTGC 156
QY 3721 GTGACCGCGGAGTGGGGCCAAACATCTCCAGGGGACCATGAAGACGAC 3770
Db 155 ACGATGTCGACGCGCGCGCGGCGGCTCGAGGAACTCGCGCGAGGTC 106

RESULT 8
US-10-156-761-5431/c
; Sequence 5431, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109

```
; SEQ ID NO 5431
; LENGTH: 1389
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1389)
US-10-156-761-5431

Query Match          0.9%; Score 49.8; DB 14; Length 1389;
Best Local Similarity 46.2%; Pred. No. 0.001;
Matches 165; Conservative 0; Mismatches 192; Indels 0; Gaps 0;

QY 3364 TCCACCTCGGGAGAGACGAGATAGAGCCCAAGGTCAAGCCGACGGCTTGTGTTGGGC 3423
DB 1313 TCCTCGTGTCCAGGTCAAGATCTCGCGGTGTGTTGCTTCAACGTTGGCGGCTCGACC 1254
QY 3424 ATCAACCTTTTCCCTGTAGCCAGAGGACGACGATCCAGGGTGGCGGATTCCTGTG 3483
DB 1253 ACAGCGGCACGGCTTCCCGCGGGGACCTCGCGGACGATCGGGACCTGGCCACCC 1194
QY 3484 TCCGTGTAGGCTTGAACCAATTGAAGGAGAGACGCGGTTGGCCATGTCTAGGGCGAGG 3543
DB 1193 TCGACGACCGCTGGCGCACTGTGTGCGGGGACATGCGGAGGAAGCCGACGCGGCATG 1134
QY 3544 TTCCTCAGGGGTAGATGATGCTCTCCAGGTGAGCCTGCTTCCCTGGCCTCAGCGG 3603
DB 1133 AGGTGCTCGGTGAGCTCTCTCTGTGTGACGTGAGCTTCTCTTGTGACGAGCTCGTCG 1074
QY 3604 ACGGCCCAAGAGGTGCCGATGCCCGGCTTCCCGGTTGACGGTTTGGTGAAGGACGTCG 3663
DB 1073 AGGACGAACTGGTCTTATGATCCCTTGACCGCGCTCTCTTGGTCTCGGCTGAACTCC 1014
QY 3664 GTGGCCACAGCGCCCTTTTCTCAAGACCTTCTCAGGGGTGAGCGTCTGCGG 3720
DB 1013 TCGACCGCTTTCCTCGATCTCAGGTAATCTCGAGGTGCGAGGCCCATCTGCGCG 957
```

RESULT 9

```
US-10-282-122A-26289/c
; Sequence 26289, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIORITY FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
```

```
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26289
; LENGTH: 1674
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
US-10-282-122A-26289

Query Match          0.8%; Score 49.6; DB 12; Length 1674;
Best Local Similarity 45.7%; Pred. No. 0.0013;
Matches 212; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

QY 3304 TCTAGGAGATCGGCATGTACGTAGCAATGAGGGTGTAAAGGCGGGGAGTTTGAACGC 3363
DB 644 TCTGTGGTGAACCGCCAGGATCGACCGGGTAGCTGGCCAGATGCTGTTTCGAGCCACTGC 585
QY 3364 TCCACCTCGGGAGGACCAAGGATGAGGCGCCAGGTCAACGGCCACCGTCTTGTGTGGGC 3423
DB 584 ACACCTTTCGCGCTTAGTGTGTTGTCGGCTCTGTCAGCAACAACAGGTTCGGGTTGGAC 525
QY 3424 ATCACCTTTTCCCTGAGCCAGGACCAAGGACCAAGTCCAGGGTGGGCGGATTCGGTGG 3483
DB 524 AACAGCAGTTTGCACAGCGCCACCGCGGCACCTCGCCACCGATAGTTGTTTACCG- 466
QY 3484 TCTGTGTAGGCTTGTACCCAGTTGAAGGAGAGACAGCCGCTTGGCCATGCTTAGGGCAGG 3543
DB 465 --CTGCTCGCGCGGACAGCGCGCATCATGCGCTGCTCGAGCTGCGCTCGAGG 408
QY 3544 TTCTCCAGGGTAGATGTAGTGTCTCAGGTGAGCTGCGCTTCCCTGGCCTCAGCCG 3603
DB 407 TCCACGCGTGGCGGTGGTCCAGTTCTCTTTCGAGCCGACCCATCTTTCATCAGCTCG 348
QY 3604 ACGGCCCAAGGTTGCGGATGGCGCGCTCCCGCTTCAACGTTTGTGTAAGGACGTCG 3663
DB 347 TCGGTGTAGTGTGGCCATCAATTCGGGCACTCGTTGAAGCGGTGAGCTTGATCTTG 288
QY 3664 GTGGCCACCAAGCCCTTTTCTCAAGGACCTTCTTCCAGGCGTGGACGTCGTCGGGGTG 3723
DB 287 ATGTCCCGCATGCGCTCTTCCACATTCGCGCAAGGCTCTTGTCTCTGTTCAAGCGCGT 228
QY 3724 ACCCCAGGTTGAGGGCGCAACATCTCCAGGGGACCATCAAGAC 3767
DB 227 TCCTGTTCAGGATGCCCAACGTTGCGCGTGGCGCGGTTGGTGGCGAAGGC 184
```

RESULT 10

```
US-10-282-122A-28540/c
; Sequence 28540, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
```



```

RESULT 15
US-10-200-562-190/c
; Sequence 190, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowan, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C2
; CURRENT APPLICATION NUMBER: US/10/200,562
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 190
; LENGTH: 9369
; TYPE: DNA
; ORGANISM: HSV2
US-10-200-562-190

Query Match      0.8%; Score 48.6; DB 14; Length 9369;
Best Local Similarity 43.9%; Pred.No. 0.0073;
Matches 207; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

QY 3333 GAGGGTGATAAGGCGCGGAGTTTGAACGCTCCACCTCGGGGAGGACCCAGGATGAGGCC 3392
Db      |||
QY 4113 GCCTTTGCGGAGTCCCGGGGTTGTAGCGGTGCGTCCCGCAGCGCGCTGCGAGCGGCG 4054
Db      |||
QY 3393 CAGGTCAACGGCCACCGTCTTGGTGTGGGCATCACCTTTTCCCTGAGCCCCAGAGGAC 3452
Db      |||
QY 4053 CAGCTCGAACCATCGAACTCGGCGGGTTCGACGCGGTCCAGCGCGCTCGACGCC 3994
Db      |||
QY 3453 CAGCAGCTCCAGGGTGGGCGGATTCGCTGCTCTGTTAGGCTTACCCAGTTGAAGGA 3512
Db      |||
QY 3993 GCGGGCCAGCGCTCGCTGCTGCCCGCGGCGGTGGGCGGCATCTTCGCCGTGAGTC 3934
Db      |||
QY 3513 GAGCAGCGGTTGGCCATGTCTAGGCGAGGTTCTTCAGGGGTAGATGATGTCCTCCAG 3572
Db      |||
QY 3933 GCGGACGGGCGCTCAAGTTCTGCGGCGGGGTGCGGTGCGGCGCGATGACCTTCCCGAG 3874
Db      |||
QY 3573 GGTGAGCTGGCTTCCCTGGCTCAGCGGACGCGCCCAAGGGTCCGATGGCCCGGCG 3632
Db      |||
QY 3873 CTCCTGAGGGCGCGCCGCTGGGGGAATGTTCCCGGCCCTCCCTTCGGCGTGAGCAG 3814
Db      |||
QY 3633 CTCCTGTTGACGGTTTGGTGAAGGACGTCTGGTGGCCACAGGCCCTTTTCTCAAGGAC 3692
Db      |||
QY 3813 GCGCCCGAACCCAGCCTCGTCCCGCCGAGGCTTTCCCGAGCAGCGGTCTGCGCGGCGC 3754
Db      |||
QY 3693 CTTCTTCCAGGCGTGACCGTCTGCGGGTGACCCCGAGTTGAGGGCCACATCTCCAG 3752
Db      |||
QY 3753 CCGGGCATGATGAGGCGGCGATGTCTCCCTCCGCGTGGCGAGGCCCGCGCGCCTG 3694
Db      |||
QY 3753 GGGGACCATGAGACGACCGTCCCGACCTTCTCCAGAGCTCCCGGTTGCC 3803
Db      |||
QY 3693 GACTACAGGTGCGGCGCGCGGACCCAGGCTGCTGAGCTCGTCGATGCC 3643
Db      |||

```

Search completed: March 18, 2004, 10:21:56
Job time : 1368.06 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 21:14:07 ; Search time 9459.64 Seconds
(without alignments)
18464.115 Million cell updates/sec

Title: US-09-664-186-6
Perfect score: 5849
Sequence: 1 tctagaaggtcagggtggac.....tgtagaagaccagaagacc.5849

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_estba:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	60	1.0	683	14	CA235176
C 2	58.6	1.0	997	29	CNS005TE
C 3	54.6	0.9	925	29	CNS0091P
C 4	53.2	0.9	655	29	CG352167

5	53.2	0.9	679	29	CC609109
6	53.2	0.9	773	29	CC661222
7	53.2	0.9	807	28	BZ529007
8	53.2	0.9	815	29	CG034209
9	53.2	0.9	897	29	CC661216
10	52.4	0.9	1201	13	BX381961
11	51.8	0.9	925	29	CNS0091P
12	51.6	0.9	437	14	CA552474
13	51.2	0.9	649	14	CA230834
14	51.2	0.9	649	14	CB870888
15	50.6	0.9	594	14	CB874255
16	50.6	0.9	637	12	BI960118
17	50.6	0.9	664	13	B0762267
18	50.6	0.9	973	14	CA275829
19	50.4	0.9	855	29	CG041883
20	50.4	0.9	890	29	CC731258
21	50.4	0.9	937	29	CC729649
22	50.4	0.9	1201	13	BX376097
23	50.2	0.9	509	12	BJ549094
24	49.8	0.9	608	14	CA179624
25	49.8	0.9	659	14	CA194749
26	49.8	0.9	694	14	CA196087
27	49.8	0.9	1201	13	BX381961
28	49.4	0.8	775	29	CG207191
29	49.4	0.8	793	29	CG449321
30	49.4	0.8	837	29	CC701337
31	49.2	0.8	502	12	EM335757
32	49.2	0.8	728	14	CF638161
33	49.2	0.8	759	10	BF259495
34	49.2	0.8	745	29	CG633883
35	49	0.8	747	29	CG633878
36	49	0.8	869	29	CG284200
37	49	0.8	909	29	CG228851
38	49	0.8	1201	13	BX376097
39	48.6	0.8	559	14	CA208502
40	48.6	0.8	564	14	CA210811
41	48.6	0.8	605	14	CA181598
42	48.6	0.8	620	14	CA182549
43	48.6	0.8	647	14	CA216848
44	48.6	0.8	730	14	CA195888
45	48.6	0.8			

ALIGNMENTS

RESULT 1
CA235176/c
LOCUS
DEFINITION
5', mRNA sequence.
CA235176
CA235176.1 GI:35303456
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CA235176 683 bp mRNA linear EST 25-SEP-2003
SCCCFL4093A04.g FL4 Saccharum officinarum cDNA clone SCCFL4093A04
5', mRNA sequence.

CA235176

CA235176.1 GI:35303456

VERSION

KEYWORDS

SOURCE

ORGANISM

Saccharum officinarum

Saccharum officinarum

EST

1 (bases 1 to 683)

Vettore A.L., da Silva, P.R., Kemper, E.L. and Arruda, P.

The libraries that made SUCEST

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

Contact: Arruda P

Centro de Biologia Molecular e Engenharia Genetica

Universidade Estadual de Campinas

Caixa Postal 6010, 13083-970, Campinas SP, Brazil

Tel: 55 19 3788 1137

Fax: 55 19 3788 1089

Email: parruda@unicamp.br

Clone distribution: clone distribution information can be found

through the Brazilian Clone Collection Center (BCCC) at

http://www.bccc.unesp.br

Plate: 093 row: A column: 04
Seq primer: T7 Promoter Primer.
Location/Qualifiers
1. 683
/organism="Saccharum officinarum"
/mol_type="rRNA"
/db_xref="taxon:4547"
/clone="SCCFL4093A04"
/lab_host="DH10B"
/clone_lib="FL4"
/notes="Organ: Developed inflorescence and rachis (20cm-long); Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from [Developed inflorescence and rachis (20cm-long)]. cDNA was prepared from polyA+ mRNA using SuperScript Plasmid System Kit (Invitrogen). The double-strand cDNAs were fractionated in a separase CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at
http://sucest.lad.ic.unicamp.br/public"

Query Match 1.0%; Score 60; DB 14; Length 683;
Best Local Similarity 46.7%; Pred. No. 0.0067;
Matches 189; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

QY 3397 TCAACGGCCACGGTCTGTGGTCATCCCTTTCCCTGAGCCGAGGACCCAGC 3456
Db 658 TGAAGGGCCAGGCCAGGTGGGNCCTTCCTTCCCTCCGACTTGGCGGTGGGC 599
QY 3457 ACCTCAGAGGTGGGGCGGATTCGGTGTCTCTGTAGCCCTTCAACCAAGTTGAAGAGAGC 3516
Db 598 GCATCGGTGGCGCGGCATCCCTGTGTCAACGTAGACGTAGTACTGTGAGCAGCAAGCG 539
QY 3517 ACCTCGTGGCGATGCTAGGGGAGGTCTCCAGGGTAGATGATGCTCCAGGGTG 3576
Db 538 AGCGGAGCGCGACCCACTCGAGGAGAGATGGGACCGCCGCGGACCATCGG 479
QY 3577 AGCTCGCTTTCCCTGGCTCGAGCGCGGCGCCAAAGGGTCCGATGCCCGCGCTCC 3636
Db 478 AGATGACGGCGCGTCTCTCGCGAGCTAGACCTGAGCCGCTCGGAGGAGACTCC 419
QY 3637 CCGTTGACGGTTGGTGAAGAGCTCGGTGGCCACAGCCCTTTTCTCAGGACCTTC 3696
Db 418 GCGTTCGGGAGAAGCGAGCGCGCGCGCGCTTCGAGGATGGAGACGACGACGCTG 359
QY 3697 TTCAGGCGTGAAGCTCTGCGGGGTGACCCGAGTTGAGGGCCAAATCTCCAGGGGG 3756
Db 358 GCGCATGTGACGTGTAGACGCGCCCGCGCGGTTCGGCGCGCGATGACGCG 299
QY 3757 ACCATGAAGACGACCTCCCGACCTTCCTCCAGAGTCCCGGTG 3801
Db 298 CACCCGCGCACCGCGCGCGCCACGGTCAAGCGGTGAGCAGGATG 254

RESULT 2
CNS005TE/c
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
AL060767
AL060767.1 GI:4943573
GSS.
Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 997)
Drosophila melanogaster genome survey sequence TET3 end of BAC #
Direct Submission

Accession
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE

JOURNAL

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrsf@genoscope.cns.fr
- Web : www.genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org/TheBDGP/Drosophila_melanogaster_BAC_library was prepared by Kazutoyo Osogawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. 997
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR12K22"
/clone_lib="RPCI-98"
/note="end : TET3"

ORIGIN

Query Match 1.0%; Score 58.6; DB 29; Length 997;
Best Local Similarity 18.7%; Pred. No. 0.018;
Matches 98; Conservative 180; Mismatches 245; Indels 0; Gaps 0;

QY 905 GAAAAACAGCAAAAGCGCCCTTGAGGAGCATTCGAGGCTTGCTCAAG 964
Db 989 GAAARARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 930
QY 965 CACCAAGAACCATCCGCTCTCGCCAGGAGCCCAAGAGAGCGCTAGAGCTTG 1024
Db 929 ARRR 870
QY 1025 ATCCAAAGGCTCAAGGAGCTCCCGAAATAATCGGCAGATAGGCGCGATGTTCAAAG 1084
Db 869 AARARGARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 810
QY 1085 TGTACAAAGAGACTAAAGGGGAAAATAGAGAGAGGCTTCAGGCCCTCAACCA 1144
Db 809 RAARARARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 750
QY 1145 AAGATTGTGTAGTATCCCTCGAAAGTAAACCGGAGCAAGCACCTTATTGGGAG 1204
Db 749 RAARRR 690
QY 1205 AGAAGACGGGCGATCATATACCGGGATCGATGAAGCTTTGAAGATCGCCCAAG 1264
Db 689 GRGAGARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 630
QY 1265 GAAACCTGGCGCTTGGCAGGAGCAGACTACCGACCAAGCGCGCTAGATTCTACGTG 1324
Db 629 GRGRR 570
QY 1325 GTCATCGCGGTAGCCCTGAAGAGACATGCACCTTAACAGGAGAGTGAAGTTCAATCC 1384
Db 569 RARRAARGAARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 510
QY 1385 GACTTTGGCGGAAACCAAGCACCCAGAACTAGTAGCAAGG 1427
Db 509 GARR 467

RESULT 3

CNS0091P

LOCUS

DEFINITION

CNS0091P 925 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #

RESULT 4	CG352167/c	655 bp	DNA	linear	GSS 26-AUG-2003
LOCUS	CG352167	655 bp	DNA	linear	GSS 26-AUG-2003
DEFINITION	OGLBWI2TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBma0727A23, genomic survey sequence.				
ACCESSION	CG352167				
VERSION	CG352167.1	GI:34269433			
KEYWORDS	GSS.				
SOURCE	Zea mays				
ORGANISM	Zea mays				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.				
AUTHORS	1 (bases 1 to 655) Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T., Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T., Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N. Consortium for Maize Genomics Unpublished (2002)				
TITLE	Other_GSSs: OGLBWI2TH				
JOURNAL	Contact: Cathy Whitelaw				
COMMENT	TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA Tel: 301-838-5843 Fax: 301-838-0208 Email: whitelaw@tigr.org Seq primer: TP Class: sheared ends.				
FEATURES	Location/Qualifiers				
source	1..655 /organism="Zea mays" /mol_type="genomic DNA" /strain="B73" /db_xref="taxon:4577" /clone="ZMMBma0727A23" /clone_lib="ZM_0.7_1.5_KB" /notes="vector: pBCSK-; Site 1: HindIII; 0.7-1.5 kb methylation filtered genomic DNA library"				
ORIGIN	Query Match 0.9%; Score 53.2; DB 29; Length 655; Best Local Similarity 46.1%; Pred. No. 0.29; Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;				
QY	3452	CCAGCAGCTCCAGGCTGGGGCGGATTCGGTGGTCCTGGTAGGCTTGACCCAGTTGAAGG	3511		
DB	530	CCAGCAGCTGCATCGTGGAGCGCGACTCCACGGGGTAGAAGCAGCGCCACCGTTGCGGT	471		
QY	3512	AGAGCAGCGCTGGGCATCTCTAGGCGGAGGTTCTCCAGGGGTAGATGATGTCGTCCA	3571		
DB	470	CCACCATGGCCAGCGCCAGAAACCCGGAAGACAGCAGCCCGTGCATGACGTGCGCA	411		
QY	3572	GGGTGAGCTTGGCTTTCCCTGGGCTCAGCCGAGCGGCCAAAGGTCGATGCCCCCGC	3631		
DB	410	CGGTGAGCTGTACCTGTCTGTCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGAGCTGG	351		
QY	3632	GCTCCCGTTGACGGTTTGGTGAAGGACGTCGGTGGCCACAGGCCCTTTTCTCAAGA	3691		
DB	350	AGCGGCCACCTCGATGGGAGCAGGCGCGGATGGTGCATGAAGCCGTACCGCAGC	291		
QY	3692	CTTCTCTCAGGCGTGACGGTCTGCGGGTGACCCCGAGTTGAGGGCCAACTCTCCA	3751		
DB	290	CGCGCCAGGCGGTGTCTCGCGTAGTGTCTGGTGAAGCAGAGGACGAAGCAGAGCGCGC	231		
QY	3752	GGGGGACATGAGACGACCGTCCCACTTCCTCCAGAGCTCCCGGTGTCGGTAGGGGA	3811		
DB	230	AGGCGCGATGAGCAGCGCCGTGAGCGCGGTTCCGCGGCTTGCAGGTACCGTACGCG	171		
QY	3812	TGTTGAGCGGGCAATCTCTCGAGG	3837		
DB	170	TGAAGGACGGCAGCAGCAATTCGAAG	145		


```
RESULT 7
BZ529007/c
LOCUS
DEFINITION
  BZ529007 807 bp DNA linear GSS 16-DEC-2002
  OGAIR91TC ZM2 0.7-1.5-KB Zea mays genomic clone ZM2M0070014,
  genomic survey sequence.
ACCESSION
  BZ529007
VERSION
  BZ529007.1 GI:27070951
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
  1 (bases 1 to 807)
  Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
  Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
  Citek,R.W., Nurnberg,A., Robbins,D. and Lakey,N.
  Consortium for Maize Genomics
  Unpublished (2002)
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TF
  Class: sheared ends.
  Location/Qualifiers
    1..807
    /organism="Zea mays"
    /mol_type="genomic DNA"
    /strain="B73"
    /db_xref="taxon:4577"
    /clone="ZM2M0070014"
    /clone_lib="ZM2 0.7-1.5 KB"
    /note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
    methylation filtered genomic DNA library"
ORIGIN
  Query Match 0.9%; Score 53.2; DB 28; Length 807;
  Best Local Similarity 46.1%; Pred. No. 0.32;
  Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
  QY 3452 CCAGCAGCTCCAGGTTGGCGGATTCGTTGCTGCTAGGCGAGGTTCTCCAGGGGTAGATGTAGTGTCCA 3571
  Db 527 CCAGCAGCTGCTGTTGACGCGGACTCCACGGGTAGAACACGCCACACGTTGCGGT 468
  QY 3512 AGAGCACGCGTTGGCCATGTCTAGGCGAGGTTCTCCAGGGGTAGATGTAGTGTCCA 3571
  Db 467 CCACCATGGCCAGCCAGGACACCGCGAACACAGCAGCCGTCATGACGTCGCGCA 408
  QY 3572 GGGTGAGCTTGCTTTCCCTGCGCTCAGCGGACGCGCCCAAGGTGCGGATGCCCGGC 3631
  Db 407 CGGTGAGCTGTACCTGTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 348
  QY 3632 GCTCCCGTTGACGTTTGGTGAAGACGTCGTTGGCGCCACAGGCGCTTTTCTCAAGGA 3691
  Db 347 AGCCGCCACCTCTGATGGGACAGCGCGCTGCTGGGTGATGAAGCGGTACCGCAGC 288
  QY 3692 CTTCTTTCAGCGCTGACGCTGTCCCGGTGACCCCGGTTGAGGGCCCAACATCTCCA 3751
  Db 287 CGCCCGTGGCGGTGTCGCGGTAGCTGTGCGTGAAGCAGAGGACGAGCAGCAGCGCGC 228
  QY 3752 GGGGACCATGAGACGACCGTCCACCTTCTCCAGAGCTCCCGGTTGCCGTAGGGGA 3811
  Db 227 AGCGCCGATGAGGACGCGCTGAGCGCGCGGTTCGCGCGTTGAGGTACCGTACGCGG 168
  QY 3812 TGGTGAGCGCGCAATCTCTCTGAGG 3837
  Db 167 TGAAGACGGCAGACCATTTCCGAG 142

RESULT 8
CG034209/c
LOCUS
DEFINITION
  CG034209 815 bp DNA linear GSS 19-AUG-2003
  PUJGY50TD ZM 0.6-1.0-KB Zea mays genomic clone ZM2M0666103,
  genomic survey sequence.
ACCESSION
  CG034209
VERSION
  CG034209.1 GI:33906365
KEYWORDS
  GSS.
SOURCE
  Zea mays
  ORGANISM
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
    clade; Panicoideae; Andropogoneae; Zea.
  1 (bases 1 to 815)
  Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
  Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
  Bennetzen,J.
  Maize Genomics Consortium
  Unpublished (2003)
  Contact: Cathy Whitelaw
  TIGR
  9712 Medical Center Drive, Rockville, MD 20850, USA
  Tel: 301-838-5843
  Fax: 301-838-0208
  Email: whitelaw@tigr.org
  Seq primer: TF
  Class: sheared ends.
  Location/Qualifiers
    1..815
    /organism="Zea mays"
    /mol_type="genomic DNA"
    /strain="B73"
    /db_xref="taxon:4577"
    /clone="ZM2M0666103"
    /clone_lib="ZM 0.6-1.0 KB"
    /note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
    Cot selected genomic DNA library"
ORIGIN
  Query Match 0.9%; Score 53.2; DB 29; Length 815;
  Best Local Similarity 46.1%; Pred. No. 0.33;
  Matches 178; Conservative 0; Mismatches 208; Indels 0; Gaps 0;
  QY 3452 CCAGCAGCTCCAGGTTGGCGGATTCGTTGCTGCTAGGCGGTTCTCCAGGGGTAGATGTAGTGTCCA 3571
  Db 523 CCAGCAGCTGCTGTTGACGCGGACTCCACGGGTAGAACACGCCACACGTTGCGGT 464
  QY 3512 AGAGCACGCGTTGGCCATGTCTAGGCGAGGTTCTCCAGGGGTAGATGTAGTGTCCA 3571
  Db 463 CCACCATGGCCAGCGCCAGAACACCGCGAACACAGCAGCAGCCGTCATGACGTCGCGCA 404
  QY 3572 GGGTGAGCTTGCTTTCCCTGCGCTCAGCGGACGCGCCCAAGGTGCGGATGCCCGGC 3631
  Db 403 CGGTGAGCTGTACCTGTCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 344
  QY 3632 GCTCCCGTTGACGTTTGGTGAAGACGTCGTTGGCGCCACAGGCGCTTTTCTCAAGGA 3691
  Db 343 AGCGCCACCTCTGATGGGAGCAGCGCGCTGCTGGGTGATGAAGCGGTACCGCAGCG 284
  QY 3692 CTTCTTTCAGGCGGTGACGCTCTGCGGGTGAACCCCGGTTGAGGGCCCAACATCTCCA 3751
  Db 283 CGCCCGTGGCGGTGTCGCGGTAGCTGTGCGTGAAGCAGAGCAGAGCAGAGCGCGC 224
  QY 3752 GGGGACCATGAAGACGACCGTCCCGACCTTCTCCAGAGCTCCCGGTTGCCGTAGGGGA 3811
  Db 223 AGCGCCGATGAGCAGCGCGCTGAGCGCGCGGTTCGCGCGTTGAGGTACCGTACCGCG 164
  QY 3812 TGGTGAGCGCGCAATCTCTCTGAGG 3837
  Db 163 TGAAGACGGCAGACCATTTCCGAG 138

RESULT 9
```


Barley ESTs from coleoptile tissue
Unpublished (2003)
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 08466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 594 Std Error: 0.00
Plate: 15 row: J column: 4
Seq primer: SP6.
Location/Qualifiers

TITLE
JOURNAL
COMMENT

FEATURES
source
1..594
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="Sloop"
/db_xref="GABI:552657"
/db_xref="taxon:4513"
/clone="HC15J04"
/tissue_type="coleoptile"
/dev_stage="coleoptile, 3 days old"
/lab_host="DH10B"
/clone_lib="CH"
/note="Vector: pSPORT; Site 1: SalI (5-end of cDNA);
Site 2: NotI (3-end of cDNA); Due to the cloning system
used blue/white selection for recombinants is not 100%
reliable. Average insert size is 1.3 kb."

ORIGIN

Query Match 0.9%; Score 50.6; DB 14; Length 594;
Best Local Similarity 46.3%; Pred. No. 1.2;
Matches 167; Conservative 0; Mismatches 194; Indels 0; Gaps 0;
QY 3351 GAGTTTGAACGCTCCACCTCGGGGAGACACGAGATGAGGCCAGGTCACGGCCACGGT 3410
Db 228 GATCTCGACGCGCTGCTCGCGGAGGCGCTTGTGAGGTCGAGACGAAGTCGCGGTA 287
QY 3411 CTTGTTGGGATCACCCTTTCCCTGAGCCAGAGACCCAGCACGTCACGAGCGTGGG 3470
Db 288 GTAGCGATGATCTTCCGATGAGGCGCTCGCGGGAGACACGAGAGCCGTTCCAGAG 347
QY 3471 GCGGATTCCGTGCTCTGTTAGGCTTGACCCAGTTGAAGGAGAGACGCGCTTGGCCAT 3530
Db 348 GGTGTTCTCTAGCCGAGGAGGCGCGGACCTTGACGAGCTTCATGAGGCGCTCGTGTA 407
QY 3531 GTCTAGGCGAGGTTCTCTCAGGGTAGATGTAGTGTCCAGGTCAGGTCGCTGCTTCC 3590
Db 408 GTTGAAGTAGTTGCTCTTGTGCGGCTCCAGAGACGAAGTCGAAGTCCGCTGCTGGC 467
QY 3591 TGGCCTCAGCCGAGCGCCCAAGGGTGCAGATGGCCCGGCGCTCCCGTTGACGGTTG 3650
Db 468 CTGCTCTCAGGAGGCGCTCCAGGAGCGCGGCGGCGGCGCTCCGGAAGTCGATCTT 527
QY 3651 GTGAAGGACGTCGGTGGCCACAGGCGCTTTTTCAGGACCTTTTTCAGGCGTGGAC 3710
Db 528 GTGCGCCACGCGCGCTTCTCGATGACAGCGGCGGCGGCGGCTAGTCTTCGCGGTTGAT 587
QY 3711 G 3711
Db 588 G 588

Search completed: March 18, 2004, 05:45:37
Job time : 9469.64 secs

Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 649 Std Error: 0.00
Plate: 15 row: J column: 4
Seq primer: T7.
Location/Qualifiers

FEATURES
source
1..649
/organism="Hordeum vulgare"
/mol_type="mRNA"
/cultivar="Sloop"
/db_xref="GABI:549289"
/db_xref="taxon:4513"
/clone="HC15J04"
/tissue_type="coleoptile"
/dev_stage="coleoptile, 3 days old"
/lab_host="DH10B"
/clone_lib="CH"
/note="Vector: pSPORT; Site 1: SalI (5-end of cDNA);
Site 2: NotI (3-end of cDNA); Due to the cloning system
used blue/white selection for recombinants is not 100%
reliable. Average insert size is 1.3 kb."

ORIGIN

Query Match 0.9%; Score 51.2; DB 14; Length 649;
Best Local Similarity 46.3%; Pred. No. 0.87;
Matches 167; Conservative 0; Mismatches 194; Indels 0; Gaps 0;
QY 3351 GAGTTTGAACGCTCCACCTCGGGGAGACACGAGATGAGGCCAGGTCACGGCCACGGT 3410
Db 587 GATCTCGACGCGCTGCTCGCGGAGGCGCTTGTGAGGTCGAGGAGTCCGCGTA 528
QY 3411 CTTGTTGGGATCACCCTTTCCCTGAGCCAGAGACCCAGCACGTCACGAGGTTGGG 3470
Db 527 GTAGCGATGATCTTCCGATGAGGCGCTCGCGGGAGACGAGCGGCTTCCAGAG 468
QY 3471 GCGGATTCCGTGCTCTGTTAGGCTTGACCCAGTTGAGGAGACGCGCGCTTGGCCAT 3530
Db 467 GGTGTTGTCGATGCGGAGGCGCGCCGACCTTGACGAGCTTCATGAGGCGCTCGTGTA 408
QY 3531 GTCTAGGCGAGGTTCTCCAGGGTAGATGTAGTCCAGGGTGAAGCTTGGCTTTCCC 3590
Db 407 GTTGAAGTAGTTGCTCTTGTGCGGCTCCAGAGACGAAGTCGAGGTCGCTGTTGGC 348
QY 3591 TGGCCTCAGCGGAGCGGCCAAGGTCGAGTGGCCCGGCGCTCCCGTTGAGCGTTG 3650
Db 347 CTCGCTCTCAGGAGGCGCTCCAGACGCGGAGCGCGGCGCTCCGCGAAGTCGATCTT 288
QY 3651 GTGAAGGACGTCGGTGGCCACAGGCGCTTTTTCAGGACCTTTTTCAGGCGTGGAC 3710
Db 287 GTGCGCCACGCGCGCTTCTCGATGACGCGGAGCGGCGGCGGCTAGTCTTCGCGGTTGAT 228
QY 3711 G 3711
Db 227 G 227

RESULT 15
CB874255
LOCUS
DEFINITION
CB874255 594 bp mRNA linear EST 03-JUL-2003
HC15J04Y CH Hordeum vulgare cDNA clone HC15J04 3-PRIME, mRNA
sequence.
ACCESSION
CB874255
VERSION
CB874255.1 GI:30076241
KEYWORDS
EST.
SOURCE
Hordeum vulgare
ORGANISM
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 594)
REFERENCE
Varshney,R.K., Zhang,H., Burton,R., Stein,N., Langridge,P. and
Graner,A.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 20:21:47 ; Search time 2608.35 Seconds
(without alignments)
17049.054 Million cell updates/sec

Title: US-09-664-186-4
Perfect score: 1026
Sequence: 1 gtgaagaacgaaaaaacctt.....ccctcaaggagcctcctga 1026

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_hcg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_un.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1026	100.0	1026	6	AR139424 Sequence
2	1026	100.0	1600	1	U9376 Thermus sp.
C 3	1026	100.0	5849	1	AF013571 Thermus s
C 4	1026	100.0	5849	6	AR139426 Sequence
C 5	59.8	5.8	308050	1	AL939124 Streptomy
C 6	56.8	5.5	2000	6	AX855393 Sequence
C 7	56.8	5.5	308000	1	AL939112 Streptomy
C 8	56.4	5.5	2030	6	AX653795 Sequence
9	56.4	5.5	2031	6	AX755741 Sequence
10	55.4	5.4	2028	6	AX660328 Sequence
11	55.4	5.4	2028	6	AX755777 Sequence
12	54.8	5.3	2028	6	AX653794 Sequence
13	54.8	5.3	2060	6	AX654685 Sequence
14	54.8	5.3	2273	8	AK106696 Oryza sat
C 15	54.8	5.3	137651	8	AC130600 Oryza sat
C 16	54.8	5.3	177374	8	AC135429 Oryza sat
C 17	54.6	5.3	299050	1	SC0939119 Streptomy
C 18	53.8	5.2	299425	1	AP005037 Streptomy
C 19	53.6	5.2	303550	1	SC0939118 Streptomy
20	53.2	5.2	1755	6	AX654077 Sequence
21	53.2	5.2	1965	6	AX659936 Sequence
22	53.2	5.2	1965	6	AX755739 Sequence
23	53.2	5.2	2303	8	AK072823 Oryza sat
24	53.2	5.2	100800	8	AC105260 Oryza sat
25	53.2	5.2	142854	8	AC136216 Oryza sat
C 26	53.2	5.2	150188	8	AP005499 Oryza sat
C 27	53.2	5.2	167446	8	AP004746 Oryza sat
C 28	52.6	5.1	298450	1	SC0939107 Streptomy
C 29	52.4	5.1	113193	1	AF357202 Streptomy
30	52.4	5.1	113193	6	AX703543 Sequence
C 31	52.4	5.1	321250	1	SC0939111 Streptomy
32	52.2	5.1	3297	6	BD180129 Highly th
33	52.2	5.1	188050	1	AL646072 Ralstonia
34	51.8	5.0	19372	1	AE004443 Pseudomon
35	51.6	5.0	1815	6	AX148314 Sequence
36	51.6	5.0	1815	6	BD017435 Transgeni
37	51.6	5.0	2498	8	ZMU95953 Zea mays vl
C 38	51.6	5.0	138203	1	AY310323 Streptomy
39	51.2	5.0	815	6	E04306 DNA encodin
40	51.2	5.0	915	1	TTHSUPDIS
41	51.2	5.0	6955	1	TTHILVB
42	50.8	5.0	2283	6	BD180126 Thermus aqu
43	50.8	5.0	276800	1	SC0939115 Highly th
44	50.8	5.0	298550	1	AP005047 Streptomy
C 45	50.8	5.0	301617	1	AE016911 Chromobac

ALIGNMENTS

RESULT 1	AR139424	AR139424	1026 bp	DNA	linear	PAT 16-JUN-2001
LOCUS	Sequence	4 from patent US 6207377.				
DEFINITION	AR139424					
ACCESSION	AR139424					
VERSION	AR139424.1	GI:14481920				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 1026)					
AUTHORS	Wayne, J. and Xu, S.-Y.					
TITLE	Method for construction of thermus-E. coli shuttle vectors and					
	identification of two Thermus plasmid replication origins					
JOURNAL	Patent: US 6207377-A 4 27-MAR-2001;					

Pred. No. is the number of results predicted by chance to have a

FEATURES		Location/Qualifiers
source	1..1026	
	/organism="unknown"	
	/mol_type="unassigned DNA"	
ORIGIN		
Query Match		100.0%; Score 1026; DB 6; Length 1026;
Best Local Similarity		100.0%; Pred. No. 1.3e-179;
Matches 1026; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
QY	1	GTGAAGACGAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGAAACCCACGAC 60
Db	1	GTGAAGACGAAAAACCTTCTTTGAAGAGCTTTACGAGGCTTTAGAGAAACCCACGAC 60
QY	61	AACACCGATGCCACTAGGGGTCAGATAGGGGTACAGGACTTCTTCTTGCCACCGAC 120
Db	61	AACACCGATGCCACTAGGGGTCAGATAGGGGTACAGGACTTCTTCTTGCCACCGAC 120
QY	121	CCCCCTCCAGATGAGGTCCGAAATCCGCTCGGAAGGGTTTACATACCAAAAAGAG 180
Db	121	CCCCCTCCAGATGAGGTCCGAAATCCGCTCGGAAGGGTTTACATACCAAAAAGAG 180
QY	181	GCATTAGATTGCTTTACCGGAAGACCATGAGGCTTCTTCTCTGCTGGGCC 240
Db	181	GCATTAGATTGCTTTACCGGAAGACCATGAGGCTTCTTCTCTGCTGGGCC 240
QY	241	CCCCCTATACCAAGCTGAACCCCGTTGGGAATGTATGTCAAGCCGTCAGGACGG 300
Db	241	CCCCCTATACCAAGCTGAACCCCGTTGGGAATGTATGTCAAGCCGTCAGGACGG 300
QY	301	CTCAGAAGTCTGAACTCTCCAGGAGATGCCCGTCCACCATCCCTACGGCAAC 360
Db	301	CTCAGAAGTCTGAACTCTCCAGGAGATGCCCGTCCACCATCCCTACGGCAAC 360
QY	361	CGGAGCTCTGAGGAAGTGGGACCGTCTCTCATGTCCTCCCTGGAGATGTTGCC 420
Db	361	CGGAGCTCTGAGGAAGTGGGACCGTCTCTCATGTCCTCCCTGGAGATGTTGCC 420
QY	421	CTCAAACCTGGGGTCAACCGGAGACCGTCCACGCTGGAAGAGTCTTGAGAAAAAG 480
Db	421	CTCAAACCTGGGGTCAACCGGAGACCGTCCACGCTGGAAGAGTCTTGAGAAAAAG 480
QY	481	GGCTCTGGTGGCCACCGACGCTCTTACCAAAACGCTCAACGGGAGCGCGGCCATCGC 540
Db	481	GGCTCTGGTGGCCACCGACGCTCTTACCAAAACGCTCAACGGGAGCGCGGCCATCGC 540
QY	541	ACCTTTGGGCGCTCGGCTGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
Db	541	ACCTTTGGGCGCTCGGCTGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 600
QY	601	TACCCCTGAGGAACTCCGCTTAGACATGGCCAGCGGTGCTCTCTCAACTGGCT 660
Db	601	TACCCCTGAGGAACTCCGCTTAGACATGGCCAGCGGTGCTCTCTCAACTGGCT 660
QY	661	AAGGCTTACAGGACCAAGGATCCGCTCCGCTGAGGAGGAGGAGGAGGAGGAGGAG 720
Db	661	AAGGCTTACAGGACCAAGGATCCGCTCCGCTGAGGAGGAGGAGGAGGAGGAGGAG 720
QY	721	GGGAAAGGGTGTGTCGCAACCAAGACCGTGGCGGTGAGCTGGGCTCATCTGTGTC 780
Db	721	GGGAAAGGGTGTGTCGCAACCAAGACCGTGGCGGTGAGCTGGGCTCATCTGTGTC 780
QY	781	CTCCCCGAGTGGAGGTTTCAAACCTCCCGGCGCTTATCACCTCATGCTACGACAT 840
Db	781	CTCCCCGAGTGGAGGTTTCAAACCTCCCGGCGCTTATCACCTCATGCTACGACAT 840
QY	841	GCGATCTCTAGATGACGCTGTTCAGACGTTTCTATGAGGCTTCTGTGGGCTGTG 900
Db	841	GCGATCTCTAGATGACGCTGTTCAGACGTTTCTATGAGGCTTCTGTGGGCTGTG 900
QY	901	GCCAGGGTGAATCCCGCGCAATATCTATTGTCGCTCCCTAATCGGGTTATCGAGAT 960
Db	901	GCCAGGGTGAATCCCGCGCAATATCTATTGTCGCTCCCTAATCGGGTTATCGAGAT 960
QY	961	TACACGATGGCCATCTGACACGACGGGAGCGTACTAGTGAAGACCCCTCAAGGAGCC 1020
Db	961	TACACGATGGCCATCTGACACGACGGGAGCGTACTAGTGAAGACCCCTCAAGGAGCC 1020
QY	1021	TCCTGA 1026
Db	1021	TCCTGA 1026
RESULT 2		
U89376		
LOCUS		1600 bp DNA linear BCT 06-OCT-1997
DEFINITION		Thermus sp. YS45 replication protein (rept) gene, complete cds.
ACCESSION		U89376
VERSION		U89376.1 GI:2465516
KEYWORDS		Thermus sp. YS45
SOURCE		Thermus sp. YS45
ORGANISM		Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae; Thermus.
REFERENCE		1 (bases 1 to 1600)
AUTHORS		Wayne, J. and Xu, S.-Y.
TITLE		Identification of a thermophilic plasmid origin and its cloning within a new Thermus-E. coli shuttle vector
JOURNAL		Gene 195 (2), 321-328 (1997)
MEDLINE		97449309
PUBMED		9305778
REFERENCE		2 (bases 1 to 1600)
AUTHORS		Wayne, J. and Xu, S.-Y.
TITLE		Direct Submission
JOURNAL		Submitted (12-FEB-1997) Research, New England Biolabs, 32 Tozer Road, Beverly, MA 01915, USA
FEATURES		Location/Qualifiers
source		1..1600
		/organism="Thermus sp. YS45"
		/mol_type="genomic DNA"
		/strain="YS45"
		/db_xref="taxon:64227"
		220..225
-35_signal		/note="putative"
-10_signal		243..248
		/note="putative"
-35_signal		278..283
		/note="putative"
-10_signal		302..307
		/note="putative"
gene		332..1369
		/gene="rept"
RBS		332..336
		/gene="rept"
CDS		344..1385
		/gene="rept"
		/function="required for plasmid replication"
		/notes="Rept"
		/codon_start=1
		/transl_table=11
		/product="replication protein"
		/protein_id="A8572144.1"
		/db_xref="GI:2465517"
		/translation="MKNEKTFEELYEALEETHDNTDTRGSDRSEDFLATDPPDP GKLEALAGFTYQKEALRIALPKDHEAFLLSVGAPPIPPAPVPVNVCOVQDGPQ KILELQELARSTIPYCNELRWKGTVVWVPLEMLNLGVTQTVAHWKVLKFN GLVATVLTQVTNGERRALGTWAVLRPKARLTLDYIYPWRNLALDMANGVLSFN WKAYVDHGIKPTLDVLWAGKRWMPNTKTVADLGLILVLEVERSKLPALITLI ATYIALDLDRSRFRFYAGLLWAVARGELPAQYLFAVLMRVIRDYTDGHLTRPGAYLV KTLKEAS"
misc_feature		1158..1166
		/gene="rept"
		/note="putative; encodes DnaA box"
misc_feature		1293..1301
		/gene="rept"
		/note="putative; encodes DnaA box"

ORIGIN

Query Match 100.0%; Score 1026; DB 1; Length 1600;
Best Local Similarity 100.0%; Pred. No. 1.2e-179;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAAGAACGAAACCTCTTTGAAGAGCTTTACAGGCTTTAGAGAAACCCAGCAC 60
DB 344 GTGAAGAACGAAACCTCTTTGAAGAGCTTTACAGGCTTTAGAGAAACCCAGCAC 403
QY 61 AACACCGATGCCACTAGGGGTGATAGAGGGGTGAGAGGACTTCTTCTTGGCCACCGAC 120
DB 404 AACACCGATGCCACTAGGGGTGATAGAGGGGTGAGAGGACTTCTTCTTGGCCACCGAC 463
QY 121 CCCCTCAGATGAGGTGCCAAATCGCTCCGAAAGGCTTTATACACCAAAAGAG 180
DB 464 CCCCTCAGATGAGGTGCCAAATCGCTCCGAAAGGCTTTATACACCAAAAGAG 523
QY 181 GCATCTAGGATGCTTTACCCGAGAAAGACATCAGGCTTTCCTTCTGTTGGGCC 240
DB 524 GCATCTAGGATGCTTTACCCGAGAAAGACATCAGGCTTTCCTTCTGTTGGGCC 583
QY 241 CCCCTATACACAGCTGAACCCCGCTTGGGAATGTATGTCGAAGCGTCCAGGACGG 300
DB 584 CCCCTATACACAGCTGAACCCCGCTTGGGAATGTATGTCGAAGCGTCCAGGACGG 643
QY 301 CCTCAGAAGCTTCTGGAACCTCTCAGAGATGCGCGCTCCACATCCCTACGGCAAC 360
DB 644 CCTCAGAAGCTTCTGGAACCTCTCAGAGATGCGCGCTCCACATCCCTACGGCAAC 703
QY 361 CGGAGCTCTCGAGGAAGGTGGGACGCTCTGCTTTCATGTCGCCCTCGAGATGTTGGCC 420
DB 704 CGGAGCTCTCGAGGAAGGTGGGACGCTCTGCTTTCATGTCGCCCTCGAGATGTTGGCC 763
QY 421 CTCAACTGGGGGTCAACCGGAGACCGTCAACGCTTGGAGAGGCTCTTGAGAAAG 480
DB 764 CTCAACTGGGGGTCAACCGGAGACCGTCAACGCTTGGAGAGGCTCTTGAGAAAG 823
QY 481 GSCCTGTGGCCACCGAGCTCTTCAACAAACCGTCAACGGGGAGCGCGGCCCATCGGC 540
DB 824 GSCCTGTGGCCACCGAGCTCTTCAACAAACCGTCAACGGGGAGCGCGGCCCATCGGC 883
QY 541 ACCCTTTGGGCGCTCGGCTGAGGCCAGGAAAGCCAGGCTCAACCTGGAGCATCATC 600
DB 884 ACCCTTTGGGCGCTCGGCTGAGGCCAGGAAAGCCAGGCTCAACCTGGAGCATCATC 943
QY 601 TACCCCTGGAGAACCTCGCCCTAGACATGCCACCGCGTCTCTCTCAACTGGGTC 660
DB 944 TACCCCTGGAGAACCTCGCCCTAGACATGCCACCGCGTCTCTCTCAACTGGGTC 1003
QY 661 AAGGCTTACCGAGCACCGGAATCCGCCACCCCTCGGACGTGCTGCTCTGGGCTCAG 720
DB 1004 AAGGCTTACCGAGCACCGGAATCCGCCACCCCTCGGACGTGCTGCTCTGGGCTCAG 1063
QY 721 GGGAAAGGGTGATGCCCAACACGACCGTGGCCGCTTGAACCTGGGCTCATCTGGTC 780
DB 1064 GGGAAAGGGTGATGCCCAACACGACCGTGGCCGCTTGAACCTGGGCTCATCTGGTC 1123
QY 781 CTCCTCCGAGGTGAGCGTTCAAAACCTCCCGCCCTTATCACCTCATGCTACGTACAT 840
DB 1124 CTCCTCCGAGGTGAGCGTTCAAAACCTCCCGCCCTTATCACCTCATGCTACGTACAT 1183
QY 841 CGCGATCTCTAGATGACCGTGTGTCAGACGTTTCTATGAGGCTTCTGTGGGCTGTG 900
DB 1184 CGCGATCTCTAGATGACCGTGTGTCAGACGTTTCTATGAGGCTTCTGTGGGCTGTG 1243
QY 901 CCCGGGTGACTCCCGCGCAATCTATTGCGTCCCTAATGCGGGTTATCCGAGAT 960
DB 1244 CCCGGGTGACTCCCGCGCAATCTATTGCGTCCCTAATGCGGGTTATCCGAGAT 1303
QY 961 TACAGGATGGCCATCTGACACGACCGGAGCGGTACCTAGTGAAGACCTCAAGAGGCC 1020
DB 1304 TACAGGATGGCCATCTGACACGACCGGAGCGGTACCTAGTGAAGACCTCAAGAGGCC 1363

QY 1021 TCCTGA 1026
DB 1364 TCCTGA 1369

RESULT 3

AF013571/c

LOCUS

DEFINITION

Thermus sp. Y845 plasmid pTsp45s: Tsp45I restriction endonuclease

(tsp45IR), Tsp45I methyltransferase (tsp45IM), and replication

protein (repT) genes, complete cds.

AF013571

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

RBS

CDS

gene

CDS

AF013571 5849 bp DNA linear BCT 03-JAN-1998
Thermus sp. Y845 plasmid pTsp45s: Tsp45I restriction endonuclease
(tsp45IR), Tsp45I methyltransferase (tsp45IM), and replication
protein (repT) genes, complete cds.

AF013571

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1. 5849

/organism="Thermus sp. Y845"

/mol_type="genomic DNA"

/strain="Y845"

/db_xref="taxon:64227"

/plasmid="pTsp45s"

<602..>1612

/gene="tsp45IR"

602..606

/gene="tsp45IR"

623..1612

/gene="tsp45IR"

/note="R Tsp45I"

/codon_start=1

/transl_table=11

/product="Tsp45I restriction endonuclease"

/protein_id="AAC46043.1"

/db_xref="GI:2735919"

/translation="MAEWNVWQSRVELLEKGYLDKLLQVYKGGSSRSVPVEEEK

LRKAYKAYEGQDSPEAEATKLVEAVLNARFKVSPFNPHPYLPVLYLVSEKAKRANK

ALKEALQVASKHPETIRVLAKAQKQVGEALIQRLKEPPEINRQIGPMFKRWYKEEL

KGKIBERLPQPKTIIVVYSPKSRPEQAPLIAEREAGIIIVTGSDEALKDAKENLG

LGHEALGKGVDFVYVIRSPPEETWHLTGKVFQDFGNGDNOKLVAKASIRLDLE

KRHIGVVDGMPVSKFRGWAGLKGKETIVTSLVLLPDLIAELYQKGEALGL"

complement (<1609..>2860)

/gene="tsp45IM"

complement (1609..2850)

/gene="tsp45IM"

/note="M Tsp45I"

/codon_start=1

/transl_table=11

/product="Tsp45I methyltransferase"

/protein_id="AAC46044.1"

/db_xref="GI:2735920"

```

/translation="MSRSYVGLTRKAPLKASKTSBPSRPLVYPCKRDEKEILDQPT
QIVLRXETILFLCGNAPLEIDPIGTFFLGNGOVLRLWMLRPPGGYAGKQVLVIDP2
YGTQDFLYGGDEIDRVATVSPKNGQLGIDTIDGPQVFEFLRERLILRLMAD3P
LIFVHIDEKYGEVFKLILDEVGRNFNHIAIASNPNKFRKAFGSKOMILVYSK
TRDYWNESASYSEIEARLFFVDENGERTYTNPLHAPGETKDGPTGRWRGILPP
PRHWRYPPEKLDLDAOGLIIVMSKNGVPRKVKYARDLKGKIKLDQVWQKDPYPR
YTEKNLMDLKLIVOTGSGNEGLVLDPPFAGSGTTLIASPLLKRRSIGIDASWEAKF
TRVLDFRPLQHKFEIVAF"
2164..4500
/notes="required for plasmid replication (putative)"
complement(2856..2860)
/genes="tsp45IM"
complement(<3132..>4169)
/genes="rept"
complement(3132..4157)
/genes="rept"
/function="required for plasmid replication"
/notes="Rept"
/codon_start=1
/transl_table=11
/product="replication protein"
/protein_id="AAC46045.1"
/db_xref="GI:2735921"
/translation="MKNEKTFEELYEALEETHDNTDTRGSDRSEDFFLATDPPD
GGAENRLAKGFTYQKALRIALPEKXHEAFLLSSVAGPPIPPAEPVGVNQAQVDPQ
KLELLEIETARSTIPYRNLWRKVTVMVLEMLALNLGVTRQTVHAWKVLKX
GLVATDLVHTNGERRALGTWAVLRPGRKARLTDYIVFWRNLALDMANGVLSFN
WYKAYVDHGIPTLDVLWAAQKRVMPNTKVAVDLGLILVLPVERSKLPALITLI
ANYLALLDDRRSRRYACGLLWAVARGELPAQYLFVLMRVLRIDYTDGHLRTPGAYLV
KILKEAS"
complement(3200..3208)
/genes="rept"
/notes="putative DnaA box required for plasmid replication"
complement(3335..3343)
/genes="rept"
/notes="putative DnaA box required for plasmid replication"
complement(4165..4169)
/genes="rept"
complement(4194..4199)
/notes="putative"
complement(4218..4223)
/notes="putative"
complement(4253..4258)
/notes="putative"
complement(4276..4281)
/notes="putative"

rep_origin
RBS
/translation="MSRSYVGLTRKAPLKASKTSBPSRPLVYPCKRDEKEILDQPT
QIVLRXETILFLCGNAPLEIDPIGTFFLGNGOVLRLWMLRPPGGYAGKQVLVIDP2
YGTQDFLYGGDEIDRVATVSPKNGQLGIDTIDGPQVFEFLRERLILRLMAD3P
LIFVHIDEKYGEVFKLILDEVGRNFNHIAIASNPNKFRKAFGSKOMILVYSK
TRDYWNESASYSEIEARLFFVDENGERTYTNPLHAPGETKDGPTGRWRGILPP
PRHWRYPPEKLDLDAOGLIIVMSKNGVPRKVKYARDLKGKIKLDQVWQKDPYPR
YTEKNLMDLKLIVOTGSGNEGLVLDPPFAGSGTTLIASPLLKRRSIGIDASWEAKF
TRVLDFRPLQHKFEIVAF"
2164..4500
/notes="required for plasmid replication (putative)"
complement(2856..2860)
/genes="tsp45IM"
complement(<3132..>4169)
/genes="rept"
complement(3132..4157)
/genes="rept"
/function="required for plasmid replication"
/notes="Rept"
/codon_start=1
/transl_table=11
/product="replication protein"
/protein_id="AAC46045.1"
/db_xref="GI:2735921"
/translation="MKNEKTFEELYEALEETHDNTDTRGSDRSEDFFLATDPPD
GGAENRLAKGFTYQKALRIALPEKXHEAFLLSSVAGPPIPPAEPVGVNQAQVDPQ
KLELLEIETARSTIPYRNLWRKVTVMVLEMLALNLGVTRQTVHAWKVLKX
GLVATDLVHTNGERRALGTWAVLRPGRKARLTDYIVFWRNLALDMANGVLSFN
WYKAYVDHGIPTLDVLWAAQKRVMPNTKVAVDLGLILVLPVERSKLPALITLI
ANYLALLDDRRSRRYACGLLWAVARGELPAQYLFVLMRVLRIDYTDGHLRTPGAYLV
KILKEAS"
complement(3200..3208)
/genes="rept"
/notes="putative DnaA box required for plasmid replication"
complement(3335..3343)
/genes="rept"
/notes="putative DnaA box required for plasmid replication"
complement(4165..4169)
/genes="rept"
complement(4194..4199)
/notes="putative"
complement(4218..4223)
/notes="putative"
complement(4253..4258)
/notes="putative"
complement(4276..4281)
/notes="putative"

misc_feature
/genes="rept"
/notes="putative DnaA box required for plasmid replication"
complement(3335..3343)
/genes="rept"
/notes="putative DnaA box required for plasmid replication"
complement(4165..4169)
/genes="rept"
complement(4194..4199)
/notes="putative"
complement(4218..4223)
/notes="putative"
complement(4253..4258)
/notes="putative"
complement(4276..4281)
/notes="putative"

RBS
-10_signal
complement(4194..4199)
-35_signal
complement(4218..4223)
-10_signal
complement(4253..4258)
-35_signal
complement(4276..4281)

ORIGIN
Query Match
Best Local Similarity 100.0%; Score 1026; DB 1; Length 5849;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAAGACGAAAAACCTCTTTGAGAGCTTTACGAGGCTTTAGAGAAACCCACGAC 60
DB 4157 GTGAAGACGAAAAACCTCTTTGAGAGCTTTACGAGGCTTTAGAGAAACCCACGAC 4098
QY 61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTGCGCACCGAC 120
DB 4097 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTCTGCGCACCGAC 4038
QY 121 CCCCCTCCAGATGAGGTGCGGAAATCCCTCGGAGAGGCTTTACATACCAAAAAGAG 180
DB 4037 CCCCCTCCAGATGAGGTGCGGAAATCCCTCGGAGAGGCTTTACATACCAAAAAGAG 3978
QY 181 GCATTTAGATGCTTTACCGAGAAAGACCATGAGGCTTCTTCTCTGTTGGGCC 240
DB 3977 GCATTTAGATGCTTTACCGAGAAAGACCATGAGGCTTCTTCTCTGTTGGGCC 3918
QY 241 CCCCCTATACCAACGACTGAAACCCCGTTGGGAATGTATGTCAAGCGCTCCAGACCGG 300
DB 3917 CCCCCTATACCAACGACTGAAACCCCGTTGGGAATGTATGTCAAGCGCTCCAGACCGG 3858
QY 301 CCTCAGAACGCTTCTGGAACTCTCTCCAGGAGATTGCCCGCTCCACCATCCCTACGGCAAC 360

```

```

DB 3857 CCTCAGAAAGCTTCTGGAACTCTCCAGGAGATTGCCCGCTCCACCATCCCTACGCAAC 3798
QY 361 CGGAGCTCTGGAGGAAAGTGGGAGCGGTCTTATAGTCCCTCCCTGGAGATGTTGGCC 420
DB 3797 CGGAGCTCTGGAGGAAAGTGGGAGCGGTCTTATAGTCCCTCCCTGGAGATGTTGGCC 3738
QY 421 CTCAACTCTGGGGTCAACCGGAGACCGTCCACCGCTTGAAGAAGTCTCTTGAGAAAAAG 480
DB 3737 CTCAACTCTGGGGTCAACCGGAGACCGTCCACCGCTTGAAGAAGTCTCTTGAGAAAAAG 3678
QY 481 GGCCTGTGTCACCGAGCTCTTCAACAAACCGTCAACGGGAGGCGCGGGCATCGGC 540
DB 3677 GGCCTGTGTCACCGAGCTCTTCAACAAACCGTCAACGGGAGGCGCGGGCATCGGC 3618
QY 541 ACCCTTTGGCGCTCGGCTGAGCGGAGGAAAGCAGGCTCACCTTGGACGACTACATC 600
DB 3617 ACCCTTTGGCGCTCGGCTGAGCGGAGGAAAGCAGGCTCACCTTGGACGACTACATC 3558
QY 601 TACCCCTGGAGAACTCGCCCTAGACATGGCCAAACGGGCTGCTCTCTTCAACTCGGTC 660
DB 3557 TACCCCTGGAGAACTCGCCCTAGACATGGCCAAACGGGCTGCTCTCTTCAACTCGGTC 3498
QY 661 AAGCCTTACAGGACACGGAATCCGCCCAACCTCGTGAAGTCTGCTCTCTGGGCTCAG 720
DB 3497 AAGCCTTACAGGACACGGAATCCGCCCAACCTCGTGAAGTCTGCTCTCTGGGCTCAG 3438
QY 721 GGGAAAAAGGTGATGCCCAACACAGACCGTGGCGCTTACCTGGGCTCATCTCTGGTC 780
DB 3437 GGGAAAAAGGTGATGCCCAACACAGACCGTGGCGCTTACCTGGGCTCATCTCTGGTC 3378
QY 781 CTCGCCGAGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCTCATTTGCTAGTACATT 840
DB 3377 CTCGCCGAGTGGAGCGTTCCAAACTCCCGGCCCTTATCACCTCATTTGCTAGTACATT 3318
QY 841 GCCGATCTCTAGTACACGCTGTTCAAGACGTTTCTATGACGCTTGTCTGGGCTGTG 900
DB 3317 GCCGATCTCTAGTACACGCTGTTCAAGACGTTTCTATGACGCTTGTCTGGGCTGTG 3258
QY 901 GCCAGGGTGAACCTCCCGCGCAATATCTATTTGCGCTCTATATGGGGTTATCCGAGAT 960
DB 3257 GCCAGGGTGAACCTCCCGCGCAATATCTATTTGCGCTCTATATGGGGTTATCCGAGAT 3198
QY 961 TACACGATGCGCATCTGACACGCGGAGCGTACTAGTGAAGACCTCAAGGAGGCC 1020
DB 3197 TACACGATGCGCATCTGACACGCGGAGCGTACTAGTGAAGACCTCAAGGAGGCC 3138
QY 1021 TCCTGA 1026
DB 3137 TCCTGA 3132

RESULT 4
AR139426/c
LOCUS
DEFINITION Sequence 6 from patent US 6207377.
ACCESSION AR139426
VERSION AR139426.1 GI:14481922
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 5849)
AUTHORS
Wayne, J. and Xu, S.-Y.
TITLE
Method for construction of thermus-B. coli shuttle vectors and
JOURNAL
Identification of two Thermus plasmid replication origins
Patent: US 6207377-A 6 27-MAR-2001;
FEATURES
Location/Qualifiers
source
1..5849
/mol_type="unassigned DNA"
ORIGIN

```

Db 3137 TCCTGA 3132

|||||

SC0939124 308050 bp DNA linear BCT 11-FEB-2003

RESULT 5

SC0939124

LOCUS

DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 21/29.

ACCESSION AL939124 AL023797 AL031031 AL031035 AL031124 AL031225 AL031231

AL031260 AL034447 AL035559 AL035569 AL0355913 AL045882

AL939124.1 GI:24413886

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D., Bateman, A., Bowles, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.H., Kieser, T., Larke, J., Murphy, L., Oliver, K., O'Neill, S., Rabinowitch, E., Rajandream, M.A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrell, B.G., Parkhill, J., and Hopwood, D.A.

TITLE Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)

JOURNAL Nature 417 (6885), 141-147 (2002)

MEDLINE 21996410

PUBMED 12000953

REFERENCE 2 (bases 1 to 308050)

AUTHORS Bentley, S.D.

TITLE Direct Submission

JOURNAL Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces sequencing team, Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: sb@sanger.ac.uk

COMMENT On or before Oct 26, 2002 this sequence version replaced gi:20520664, gi:20520751, gi:20520752, gi:20520665, gi:20520811, gi:20520756, gi:20520845, gi:20520760, gi:20520817, gi:20520818, gi:20520782.

FEATURES

source

1. 308050

/organism="Streptomyces coelicolor A3(2)"

/mol_type="genomic DNA"

/strain="A3(2)"

/db_xref="taxon:100226"

250..2490

/gene="SCO5495"

/note="synonym: SC8D9.07"

250..2490

/gene="SCO5495"

/note="SC8D9.07, possible membrane associated phosphodiesterase, len: 746 aa; similar to many eg. Tr:O87376 (EMBL:AF052518) c-di-GMP phosphodiesterase A from Acetobacter xylinus cdg2 operon (752 aa) fasta scores; opt: 959, z-score: 1059.2, E(): 0, (31.7% identity in 603 aa overlap). Appears to have two distinct regions with the first 300aa containing several potential regions spanning hydrophobic regions and the remainder of the protein showing similarity to the phosphodiesterase mentioned above. Also similar to a nearby gene on the same cosmid (SC8D9.22) fasta scores: opt: 736, z-score: 625.6, E(): 5.3e-30, (32.8% identity in 729 aa overlap). Contains TTA leu codon, possible target for BldA regulation. Also contains Pfam match to entry PF00990 DUF9, Domain of unknown function, score 166.70, E-value 3.9e-46 and Pfam match to entry PF00563 DUF2, Domain of unknown function 2, score 291.30, E-value 1.2e-83. Contains possible hydrophobic membrane spanning regions"

/codon_start=1

/transl_table=11

/product="putative phosphodiesterase"

(EMBL:AF008553) Glu-tRNA-Gln amidotransferase subunit C, gatC, from *Bacillus subtilis* (101 aa) fasta scores; opt: 232. z-score: 294.8, E(): 4.1e-09, (42.1% identity in 95 aa overlap)."

/protein_id="CAB37571.1"
/db_xref="GI:4467246"
/db_xref="GOA:Q9Z584"
/db_xref="SPTREMBL:Q9Z584"
/translation="VEPTSSAAGSLRLRRNADAWHYRWLQDQGGVPSSMTGTRV
PQVGRPHGPQCHPLVDPOGERHLSNPALPAAVATAAFLVAGAYRAPSDGHAL
FVAGTVGSLVLSGVHGLVHMLGSRWGGTGSALTLVLLIYIPAGWVSLT
VVLGVARGRWQVGLGAVDILGIAAGAVLGVGVSPVSPWPDPETWGVYAAP
KVLVAVYLAIVTRALLWTIQTPGLTPAARTLVRHGLVAVALLGAPICVAVV
KPLLFPAIPLALDSTLWIAARAARAEQRLDTGLFNQWLLETTALDDAERIG
AAALMLIDLPRSVNDTLGLHAGDLLQTLADRLQALPGCAEARLGGDFAVLL
PVADSLTSTRARSLVGLSSPMDLDGTLVLEASAGVAFEDHALDAEGLRADY
AMQAKRRTGVEVYESKSDSTNPRLGLLLDLRLDALDAHEVQLHYQKVPDGOVAG
LEALVWHPKGVKPPDFPIAIAESSGUMPHLTETVLETLGQVQARWRAGLFPVFA
VNVSPEDVTPGFAGSVAARLARHGVFAGALQLEITEHVLLDPQRAADTNALRGHG
LVMSDDFTGSLVHLRLRPVSEKIDRSFVARLADNEDAEIKVRCVTDLAHSLGL
KVAEGVEDETWERLRLDLCDAVQGLVAAAMPPEETTAWLARGSRGWRAAALP
AAASDR"

misc_feature
1141..1635
/gene="SCO5495"
/notes="Fram match to entry PF00990 DUF9, Domain of unknown
function, score 166.70, E-value 3.9e-46"
1420..1422
/gene="SCO5495"
/notes="TAA Leu codon, possible target for bldA regulation"

misc_feature
1672..2406
/gene="SCO5495"
/notes="Fram match to entry PF00563 DUF2, Domain of unknown
function 2, score 291.30, E-value 1.2e-83"
2774..3145
/gene="SCO5496"
/notes="synonym: SC8D9.08"

gene
2774..3145
/gene="SCO5496"
/notes="SCO5496"
/notes="SCO8D9.08, small hydrophobic membrane protein, len:
123 aa; no significant database similarities. Contains
possible hydrophobic membrane spanning regions"

CDS
/codon_start=1
/transl_table=11
/product="putative small membrane hydrophobic protein"
/protein_id="CAB37572.1"
/db_xref="GI:4467247"
/db_xref="SPTREMBL:Q9Z583"
/translation="MTVTATYTLAVVNLFLCFLGYRFLFPQASAAATGVPADAGGDA
AAVLTVGVDRDGLGVVGLALLAFAGARPEANFMLCVLVLADTLILVRHGCTRAVA
FGIHFAITAVLISALLFAV"
3177..3671
/gene="SCO5497"
/notes="synonym: SC8D9.09"

gene
3177..3671
/gene="SCO5497"
/notes="SCO8D9.09, unknown, len: 164aa; similar to many of
undefined function eg. TR:O06336 (EMBL:Z95390)
hypothetical protein from *Mycobacterium tuberculosis* (177
aa) fasta scores; opt: 171, z-score: 213.4, E(): 0.00014,
(34.9% identity in 109 aa overlap)."

CDS
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAB37573.1"
/db_xref="GI:4467248"
/db_xref="SPTREMBL:Q9Z582"
/translation="MSLFPVKDESVVREAEVVGRAPTVRLRLASSSTGGALST
QRYTLTAGDAGKPHWDSAMFFLDGADVLSGDEVLTAGPGLVIVPGKPHAF
AAVPGADALLIVLAFGVRFERYFLQRIALGEATPESLLEVLNHDNLFUSSAWD
ARRG"

gene
3816..4112
/gene="SCO5498"
/notes="synonyms: gatC, SC8D9.10"

CDS
3816..4112
/gene="SCO5498"
/notes="SCO8D9.10, probable Glu-tRNA-Gln amidotransferase
subunit C, gatC, len: 98aa; similar to many eg. TR:O06492

Query Match 5.8%; Score 59.8; DB 1; Length 308050;
Best Local Similarity 46.3%; Pred No. 0.16; 227; Indels 0; Gaps 0;
Matches 196; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

QY 374 GGAAGTGGGACCGCTCTTCATGTGCCCTGGAGATGTTGGCCCTCAACTGGGG 433
GGGAGGCGCTGGAGCGCGAGGGCCTGCTCACCCCGCTGCTGCCCACTGGAGCGGGTCC 10802

DB 107969 GGGAGGCGCTGGAGCGCGAGGGCCTGCTCACCCCGCTGCTGCCCACTGGAGCGGGTCC 10802

QY 434 TCACCGGACACCGCTCCACGCGCTGGAGAAGGTCTTGGAGAAAAGCGCTGTCGCCCA 493
GCTCCGCGCCCGACGCGCAACCGCTGACCTGTGGACCTGTGGACCGGACCTGATTCAGA 10808

DB 108029 GCTCCGCGCCCGACGCGCAACCGCTGACCTGTGGACCTGTGGACCGGACCTGATTCAGA 10808

EAULTAVGARETVLYRAGABEMRAATASEGQDTAVERLVEHCGRRRLIIVLDNCEHV
DAAARAEELAAACPRITVLATREPLGVESLRVPEPLPEAALRLIADRAAARF
FYVDADEETAAACPRITVLATREPLGVESLRVPEPLPEAALRLIADRAAARF
RTVLPROQTUAVVVDWMDLLDADEREVLGRISVFAAGCDLAAEAACVCPAALDLSGS
LVDSLVVAAPQVLAALERIEYENLTLTALHAVERDEQALCLALSLVWYQWMDL
RTDPLLRGPRQVLAALERIEYENLTLTALHAVERDEQALCLALSLVWYQWMDL
RMEARNVFEVMAALGPDPFAPPGTFFAPVWERTSAPPMGTGVLAEAFRGVRLHRLMD
CMTDELDAWNPAAQRKLRIADTFEPGMPCTCSGGLWYFVMTLMDRILHRLMD
ATVTRCSETPGYEWELAGGLQMRANMLNLTWAGDAVRDADESLEYIDRLGDWAGWA

Query Match 5.5%; Score 56.8; DB 1; Length 300800;
Best Local Similarity 45.6%; Pred. No. 0.58; 237; Indels 0; Gaps 0;
Matches 199; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

QY 339 CTCACCATCCCTACGCGAACCAGGAGCTCTGAGGAGAGGTGGGACCGTCTCTTCAT 398
DB 171393 CTCGCGCTCGCCAGCTCGCCGACCTCTAGTTCAGCGAATCAAGCGCTGGTCTTCGG 171334

QY 399 GGTCCCTCGAGATGTGGCCCTCAACCTGGGGGTACCGGGAGACCGTCCAGCGCTG 459
DB 171393 GTTCATCGCGGCGATCGTCGCGCTTACCGGGGACTCAACCGCGCGCGGCCCAAGG 171274

QY 459 GAAGAAGTCTTTCAGAAAAAGGCGCTGTGGCCACCGACGTCCTTTCACCAAAACCGTCAA 518
DB 171273 CGTCGGGACCGCGCTCAACCACTCGTCTCATCACTTCTCTCTCTCTCTCTCTCA 171214

QY 519 CGGGAGCGCGGCGCATTCGCAACCTTTTGGGCGCTCGGCTGAGCGGAGGAAAGCCAG 578
DB 171213 CATGGTATGACGCGCGCTTACCTCCAGATCGTCCCGGAGGAGGAGCTGAGTCCGAT 171154

QY 579 GCTCAGCTGAGCAGTACATCTACCCCTGGAGGAACTCGCCCTAGACATGCGCAACGG 638
DB 171153 GGCTTCCCGCTGTCTGGCTGAGCGCTCGGGCGACCACTGCTCTTCTACGTCCGGC 171094

QY 639 CGTGTCTCTTCACTGGGTCAAGGGCTTACAGGACCAACGAAATCCGCCCCACCTTGA 698
DB 171093 CTTGCTGTGGTTCGGGCGACCTCGCGCTACTCAAGAGGTGCGAGCGCTCTCTCGC 171034

QY 699 CGTGTGTCTCTGGGTGAGGAAAGGTGATGATGCCCAACCAAGACCGTGGCGGT 758
DB 171033 CGAGGTGGCTTCGGCTCGCGCGCTCGCGCTCATCGGGGACCACTCGCGGTGATG 170974

QY 759 TGACCTGGCGCTCATC 774
DB 170973 CGCGATGAGCTCTTC 170958

RESULT 8
AX653795 2030 bp DNA linear PAT 22-MAR-2003
LOCUS
DEFINITION Sequence 3665 from Patent WO03000898.
ACCESSION AX653795
VERSION AX653795.1 GI:29156609
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1
AUTHORS Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
Katagiri,F., Qian,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 3665 03-JAN-2003;
SYNGENTA PARTICIPATIONS AG (CH)
FEATURES
source Location/Qualifiers
1..2030
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 5.5%; Score 56.4; DB 6; Length 2030;

Best Local Similarity 46.7%; Pred. No. 1.6;
Matches 214; Conservative 0; Mismatches 241; Indels 3; Gaps 1;

QY 356 GCAACCGGAGCTCTGAGGAGGTGGGACCGTCTTCTATGTGTCCTCCCTGGAGATCT 415
DB 485 GGCCATGTGTCTACCGGATGAAGGAGACGGCGAGGCTTACTTCGGCGAAGGTCA 544

QY 416 TGGCCCTCAACCTGGGGGTCAACCGGAGACCGTCCACGCTTGAAGAGGTCTTTGAGA 475
DB 545 CGCGCGCTCTGTCACCGTCCCGGCTACTTCAACGAGCGCGGAGGACCAAGG 604

QY 476 AAAAGGCGCTGTGGCCACCGAGCTCTTCAACCAACCGTCAACGGGAGCGCGGCA 535
DB 605 AGCGCGCTCATCGCGGGCTCACCGTTCGACCGCATCATCAACGAGCGCGCGCG 664

QY 536 TGGGACACCTTTTGGGCGTCCGCTGAGGCCAGGAAAGCCAGGCTCAACCTTGGAGCT 595
DB 665 CCATCGCTACGCGATCGACAGAGAGGCGCGAGAGAAAGCTCTCTGCTTTCGACCTCG 724

QY 596 ACATCTACCTTGGAGGAACCTCGCCCTAGACATGCGCAACGGGCTGTCTCTTCAACT 655
DB 725 GCGGCGGACGTTGCGACGTCAGCATCTCTCGCATCGACAAACGGGCTTTCGAGGTCT 784

QY 656 GGTCAAGCGCTTACAGGACCAACGGAATC--CGCCACACCTGAGCGTGTGCTCTCT 712
DB 785 CCACCAACGGGACACCGCTCGGGGAGGAGTTCGACCAACGCTCATGAGCACT 844

QY 713 GGGCTCAGGGGAAAAGGTGATGCCCAACCAACCGTGGCGGTGACCTTGGGCTCA 772
DB 845 TCGTCAAGTCTATCGCGGAGACGCGGCGGACATCGCGGCGACGCGCGCGCTGG 904

QY 773 TCTGTCTCTCCCGAGGTGGAGCTTCCAACTCCG 810
DB 905 GCAAGTCCCGCGAGTCCGAGCGCGCAACGCGCG 942

RESULT 9
AX755741 2031 bp DNA linear PAT 23-JUN-2003
LOCUS
DEFINITION Sequence 443 from Patent WO03000905.
ACCESSION AX755741
VERSION AX755741.1 GI:32167954
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1
AUTHORS Zhu,T., Cheng,W., Briggs,S., Cooper,B., Goff,S.A., Moughamer,T.,
Glazebrook,J., Katagiri,F., Krens,J., Provart,N. and Rieke,D.
TITLE Identification and characterization of plant genes
JOURNAL Patent: WO 03000905-A 443 03-JAN-2003;
SYNGENTA PARTICIPATIONS AG (CH)
FEATURES
source Location/Qualifiers
1..2031
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 5.5%; Score 56.4; DB 6; Length 2031;
Best Local Similarity 46.7%; Pred. No. 1.6;
Matches 214; Conservative 0; Mismatches 241; Indels 3; Gaps 1;

QY 356 GCAACCGGAGCTCTGAGGAGGTGGGACCGTCTTCTATGTGTCCTCCCTGGAGATCT 415
DB 485 GGCCATGTGTCTACCGGATGAAGGAGACGGCGAGGCTTACTTCGGCGAAGGTCA 544

QY 416 TGGCCCTCAACCTGGGGGTCAACCGGAGACCGTCCACGCTTGAAGAGGTCTTTGAGA 475
DB 545 CGCGCGCTCTGTCACCGTCCCGGCTACTTCAACGAGCGCGGAGGACCAAGG 604

QY 476 AAAAGGGCTGGTGGCCACCGAGCTCTTCAACAAACCGTCAACGGGAGCGCGGCCA 535
Db 605 AGCCGGGCTCATCGCCGGGCTCACCGTCAGCCGATCATCAACGAGCGACCGCGCGG 664
QY 536 TCGGCAACCTTTTGGGCGCTCGGCTGAGCCAGGGAAGCCAGGCTCAACCTTGGACGACT 595
Db 665 CCATCGCTACGGCATCGACAAGAGGGCGCGAGAAACGCTCTCGTCTTCGACCTCG 724
QY 596 ACATCTACCTCGGAGAACTCGCCCTAGATGCGCAACGCGGCTGCTCTCTTCAACT 655
Db 725 GCGGCGGCTACGCGGCTACCGCTCGCATCGCATCGACAAACCGGCTGCTTCGAGGCTCTTG 784
QY 656 GGTCAAGGCTACCAAGGACCAAGGAATC---CGCCCAACCTTGAACGCTGTGTCCTCT 712
Db 785 CCACCAACGGGACACCACTCGCGGCGAGGACTTCGACCAACGCTCATGGACCACT 844
QY 713 GGGCTCAGGGGAAAGGGTATGCCCAACCAAGACCGTGGCGCTTGACCTTGGGCTCA 772
Db 845 TCGTCAAGGTATCCCGCGGAAGACAGCGGCGGACATCGCCGCGACGCGCGCGCTGG 904
QY 773 TCCTGTCTCTCCCGAGGTGGAGCGTTCCAAACCTCCCG 810
Db 905 GCAAGCTCCCGCGAGTGGAGCGCGCAAGCGCGG 942

RESULT 10
AX660328
LOCUS AX660328 2028 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 695 from Patent WO03000906.
ACCESSION AX660328
VERSION AX660328.1 GI:29162163
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1
AUTHORS Glazebrook,J., Briggs,S., Cooper,B., Goff,S.A., Moughamer,T.,
Katagiri,F., Krepis,J., Provart,N., Rieke,D. and Zhu,T.
TITLE Plant disease resistance genes
JOURNAL Patent: WO 03000906-A 695 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
source
1. .2028
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 5.4%; Score 55.4; DB 6; Length 2028;
Best Local Similarity 46.5%; Pred. No. 2.5;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

QY 356 GCAACGGGAGCTCTGAGGAAGTGGGAGCGTCTTTCATGTGCCCTTGGAGATGT 415
Db 485 GCGCATGTGTCTACCCGGATGAAGAGACGCGCGGCTACCTCGCGGAGAGGTCA 544
QY 416 TGGCCCTCAACCTTGGGGGTACCCGGCAGACCGTCCACGCTTGAAGAGGTCTTGAGA 475
Db 545 CGCGCGCGTCTGCTACCGTCCGCGCTACTTCAACGACGCGGAGCGGCGGCGCAAG 604
QY 476 AAAAGGGCTGGTGGCCACCGAGCTCTTCAACAAACCGTCAACGGGAGCGCGGCCA 535
Db 605 AGCCGGGCTCATCGCGGCTCACCGTCAGCCGATCATCAACGAGCGACCGCGCGG 664
QY 536 TCGGCAACCTTTTGGGCGCTCGGCTGAGCCAGGGAAGCCAGGCTCAACCTTGGACGACT 595
Db 665 CCATCGCTACGGCATCGACAAGAGGGCGCGAGAAACGCTCTCGTCTTCGACCTCG 724
QY 596 ACATCTACCTCGGAGAACTCGCCCTAGATGCGCAACGCGGCTGCTCTCTTCAACT 655
Db 725 GCGGCGGCTACGCGGCTACCGCTCGCATCGCATCGACAAACCGGCTGCTTCGAGGCTCTTG 784
QY 656 GGTCAAGGCTACCAAGGACCAAGGAATC---CGCCCAACCTTGAACGCTGTGTCCTCT 712
Db 785 CCACCAACGGGACACCACTCGCGGCGAGGACTTCGACCAACGCTCATGGACCACT 844
QY 713 GGGCTCAGGGGAAAGGGTATGCCCAACCAAGACCGTGGCGCTTGACCTTGGGCTCA 772
Db 845 TCGTCAAGGTATCCCGCGGAAGACAGCGGCGGACATCGCCGCGACGCGCGCGCTGG 904
QY 773 TCCTGTCTCTCCCGAGGTGGAGCGTTCCAAACCTCCCG 810
Db 905 GCAAGCTCCCGCGAGTGGAGCGCGCAAGCGCGG 942

QY 656 GGGTCAAGGCTTACCAGGACCAAGGAATC---CGCCCAACCTTGAACGCTGTGTCCTCT 712
Db 785 CCACCAACGGGACACCACTCGCGGCGAGGACTTCGACCAACGCTCATGGACCACT 844
QY 713 GGGCTCAGGGGAAAGGGTATGCCCAACCAAGACCGTGGCGCTTGACCTTGGGCTCA 772
Db 845 TCGTCAAGGTATCCCGCGGAAGACAGCGGCGGACATCGCCGCGACGCGCGCGCTGG 904
QY 773 TCCTGTCTCTCCCGAGGTGGAGCGTTCCAAACCTCCCG 810
Db 905 GCAAGCTCCCGCGAGTGGAGCGCGCAAGCGCGG 942

RESULT 11
AX755777
LOCUS AX755777 2028 bp DNA linear PAT 23-JUN-2003
DEFINITION Sequence 517 from Patent WO03000905.
ACCESSION AX755777
VERSION AX755777.1 GI:32167972
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1
AUTHORS Glazebrook,J., Katagiri,F., Krepis,J., Provart,N. and Rieke,D.
TITLE Identification and Characterization of plant genes
JOURNAL Patent: WO 03000905-A 517 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
source
1. .2028
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 5.4%; Score 55.4; DB 6; Length 2028;
Best Local Similarity 46.5%; Pred. No. 2.5;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

QY 356 GCAACGGGAGCTCTGAGGAAGTGGGAGCGTCTTTCATGTGCCCTTGGAGATGT 415
Db 485 GCGCATGTGTCTACCCGGATGAAGAGACGCGCGGCTACCTCGCGGAGAGGTCA 544
QY 416 TGGCCCTCAACCTTGGGGGTACCCGGCAGACCGTCCACGCTTGAAGAGGTCTTGAGA 475
Db 545 CGCGCGCGTCTGCTACCGTCCGCGCTACTTCAACGACGCGGAGCGGCGGCGCAAG 604
QY 476 AAAAGGGCTGGTGGCCACCGAGCTCTTCAACAAACCGTCAACGGGAGCGCGGCCA 535
Db 605 AGCCGGGCTCATCGCGGCTCACCGTCAGCCGATCATCAACGAGCGACCGCGCGG 664
QY 536 TCGGCAACCTTTTGGGCGCTCGGCTGAGCCAGGGAAGCCAGGCTCAACCTTGGACGACT 595
Db 665 CCATCGCTACGGCATCGACAAGAGGGCGCGAGAAACGCTCTCGTCTTCGACCTCG 724
QY 596 ACATCTACCTCGGAGAACTCGCCCTAGATGCGCAACGCGGCTGCTCTCTTCAACT 655
Db 725 GCGGCGGCTACGCGGCTACCGCTCGCATCGCATCGACAAACCGGCTGCTTCGAGGCTCTTG 784
QY 656 GGTCAAGGCTACCAAGGACCAAGGAATC---CGCCCAACCTTGAACGCTGTGTCCTCT 712
Db 785 CCACCAACGGGACACCACTCGCGGCGAGGACTTCGACCAACGCTCATGGACCACT 844
QY 713 GGGCTCAGGGGAAAGGGTATGCCCAACCAAGACCGTGGCGCTTGACCTTGGGCTCA 772
Db 845 TCGTCAAGGTATCCCGCGGAAGACAGCGGCGGACATCGCCGCGACGCGCGCGCTGG 904
QY 773 TCCTGTCTCTCCCGAGGTGGAGCGTTCCAAACCTCCCG 810

Db 905 GCAAGCTCCGGCGGAGTGGAGCGCGCCAAAGCGCGG 942

RESULT 12
AX653794
LOCUS AX653794 2028 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 3664 from Patent WO03000898.
ACCESSION AX653794
VERSION AX653794.1 GI:29156608
KEYWORDS
SOURCE Oryza sativa

ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 3664 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
Location/Qualifiers
1..2028
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 5.3%; Score 54.8; DB 6; Length 2028;
Best Local Similarity 46.5%; Pred. No. 3.2;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;
QY 356 GCAACCGGAGCTCTGGAGGAGTGGGGAGCGTCTTCATGGTCCCTCGAGATGT 415
Db 485 CGCCATGGTCTCACCGGATGAAGAGACGGCGGAGGCTACCTCGGAGAGGTCA 544
QY 416 TGGCCCTCAACCTGGGGTCAACCGGACGACCGTCCACGCTGGAAGAGTCTTGAGA 475
Db 545 CGCGCGCGTCTGTCACCGTCCCGGCTACTTCAACGACGCGCAGCGCACCAGG 604
QY 476 AAAGGGCTGGTGGCCACCGAGCTCTTCAACAAACCGTCAACGGGGAGCGCGGCCA 535
Db 605 ACGCCGGTCTATCGCGGGTCAACCGTCAACGACGCGCAGCGCAGCGCAGCGCG 664
QY 536 TCGGACCTTTGGGCGGTCCGCTGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGG 595
Db 665 CCATCGCTACCGGATCGACAAAGAGGCGCGGAGGAGGAGGAGGAGGAGGAGGAG 724
QY 596 ACATCTACCCCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 655
Db 725 CGCGCGCTACCGATCGACAAAGAGGCGCGAGAGAGGAGGAGGAGGAGGAGGAGG 784
QY 596 ACATCTACCCCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 655
Db 725 CGCGCGCTACCGATCGACAAAGAGGCGCGAGAGAGGAGGAGGAGGAGGAGGAGG 784
QY 656 GGGTCAAGGCTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 712
Db 785 CCACCAACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 844
QY 713 GGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 772
Db 845 TCGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 904
QY 773 TCCTGGTCTCCCGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 810
Db 905 GCAAGCTCCGGCGGAGTGGAGCGCGCCAAAGCGCGG 942

RESULT 14
AX654685
LOCUS AX654685 2060 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 4555 from Patent WO03000898.
ACCESSION AX654685
VERSION AX654685.1 GI:29157499
KEYWORDS
SOURCE Oryza sativa

ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 3664 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
Location/Qualifiers
1..2028
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 5.3%; Score 54.8; DB 6; Length 2028;
Best Local Similarity 46.5%; Pred. No. 3.2;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;
QY 356 GCAACCGGAGCTCTGGAGGAGTGGGGAGCGTCTTCATGGTCCCTCGAGATGT 415
Db 485 CGCCATGGTCTCACCGGATGAAGAGACGGCGGAGGCTACCTCGGAGAGGTCA 544
QY 416 TGGCCCTCAACCTGGGGTCAACCGGACGACCGTCCACGCTGGAAGAGTCTTGAGA 475
Db 545 CGCGCGCGTCTGTCACCGTCCCGGCTACTTCAACGACGCGCAGCGCACCAGG 604
QY 476 AAAGGGCTGGTGGCCACCGAGCTCTTCAACAAACCGTCAACGGGGAGCGCGGCCA 535
Db 605 ACGCCGGTCTATCGCGGGTCAACCGTCAACGACGCGCAGCGCAGCGCAGCGCG 664
QY 536 TCGGACCTTTGGGCGGTCCGCTGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 595
Db 665 CCATCGCTACCGATCGACAAAGAGGCGCGAGAGAGGAGGAGGAGGAGGAGGAGGAG 724
QY 596 ACATCTACCCCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 655
Db 725 CGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 784
QY 656 GGGTCAAGGCTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 712
Db 785 CCACCAACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 844
QY 713 GGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 772
Db 845 TCGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 904
QY 773 TCCTGGTCTCCCGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 810
Db 905 GCAAGCTCCGGCGGAGTGGAGCGCGCCAAAGCGCGG 942

ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
AUTHORS Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
TITLE Plant genes involved in defense against pathogens
JOURNAL Patent: WO 03000898-A 4555 03-JAN-2003;
Syngenta Participations AG (CH)
FEATURES
Location/Qualifiers
1..2060
/organism="Oryza sativa"
/mol_type="unassigned DNA"
/db_xref="taxon:4530"

ORIGIN
Query Match 5.3%; Score 54.8; DB 6; Length 2060;
Best Local Similarity 46.5%; Pred. No. 3.2;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;
QY 356 GCAACCGGAGCTCTGGAGGAGTGGGGAGCGTCTTCATGGTCCCTCGAGATGT 415
Db 485 CGCCATGGTCTCACCGGATGAAGAGACGGCGGAGGCTACCTCGGCGAGAGGTCA 544
QY 416 TGGCCCTCAACCTGGGGTCAACCGGACGACCGTCCACGCTGGAAGAGTCTTGAGA 475
Db 545 CGCGCGCGTCTGTCACCGTCCCGGCTACTTCAACGACGCGCAGCGCACCAGG 604
QY 476 AAAGGGCTGGTGGCCACCGAGCTCTTCAACAAACCGTCAACGGGGAGCGCGGCCA 535
Db 605 ACGCCGGTCTATCGCGGGTCAACCGTCAACGACGCGCAGCGCAGCGCAGCGCG 664
QY 536 TCGGACCTTTGGGCGGTCCGCTGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 595
Db 665 CCATCGCTACCGATCGACAAAGAGGCGCGAGAGAGGAGGAGGAGGAGGAGGAGGAG 724
QY 596 ACATCTACCCCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 655
Db 725 CGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 784
QY 656 GGGTCAAGGCTTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 712
Db 785 CCACCAACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 844
QY 713 GGGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 772
Db 845 TCGTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 904
QY 773 TCCTGGTCTCCCGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 810
Db 905 GCAAGCTCCGGCGGAGTGGAGCGCGCCAAAGCGCGG 942

ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
AUTHORS The Rice Full-Length cDNA Consortium, National Institute of
Agricultural Sciences Rice Full-Length cDNA Project Team:
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,

REFERENCE	3	(bases 1 to 137651)
AUTHORS	Chow,T.-Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (25-NOV-2003) Institute of Botany, Academia Sinica, 128, Section 2, Academia Road, Nankang, Taipei 11529, Taiwan	
COMMENT	On Nov 25, 2003 this sequence version replaced gi:22212951.	
FEATURES	Location/Qualifiers	
source	1..137651 /organism="Oryza sativa (japonica cultivar-group)" /mol_type="genomic DNA" /cultivar="Nipponbare" /db_xref="taxon:39947" /chromosome="5" /clone="OSJNB0048I21"	
ORIGIN		
Query Match	5.3%;	Score 54.8; DB 8; Length 137651;
Best Local Similarity	46.5%;	Pred. No. 1.6;
Matches 213;	Conservative 0;	Mismatches 242; Indels 3; Gaps 1;
QY	356	GCAACCGGAGCTCTGGAGAAAGTGGGGAACGTCTTCATGTCGCCCTGGAGAATGT 415
Db	18020	GCGCCATGTGTTCAACCCTGATGAAGGACGCGCAGCGCTACCTCGCGAGAAGTCA 17961
QY	416	TGGCCCTCAAACCTGGGGGTCAACCGGCAGAACGCTCACGCTTGGAAGAGTCTTTGAGA 475
Db	17960	CAGCGCCGCTGCTGACCTGCCGGCTACTTCAACGACGCGAGCGGAGGCCACCAAGG 17901
QY	476	AAAAGGGCTGTGTGGGCCACCGACGTCCTTCAACAAACCGTCAACGGGAGCGCCGGGCCA 535
Db	17900	ACGCGCGGTCTACGCGGGCTCACGCTGACGCGATCATCAACGAGCGCACCGCCGCG 17841
QY	536	TCGGCAACCTTTTGGGCGCTCGGGCTGAGGCCAGGAAGCCAGGCTACCTCTGAGACACT 595
Db	17840	CCATCGCCTACGSCATCGACAAGAGGGCGCGAGAAGAACCTCCTCTTCGACCTCG 17781
QY	596	ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGCCCCAAGGGCGTCTCCTTCAACT 655
Db	17780	GCGGCGGAGTTCGACGTACGATCCTCGCATCGAACACGCGGTGTCGAGGTCCTTG 17721
QY	656	GGGTCAAGCGCTACACGAGCACCGGAATCCG---CCCCACCTGGAAGTGTGTPCTCT 712
Db	17720	CCACCACGCGACACCCACCTCGCGCGGAGGACTTCACCAACGCTCATGACCACACT 17661
QY	713	GGGCTCAGGGGAAAAAGGTGATGCCCAACACCAAGACCGTGCCCGTTGACCTGGSCCTCA 772
Db	17660	TGTTAAGTATCCGCCGAAGACACGGCGCGCATGCCCGGACGCGCGCGCTGG 17601
QY	773	TCTGTCTCTCCCAGGTGGAGCGTTCCAAACTCCCG 810
Db	17600	GCAAGCTCCGCCGAGTTCGAGCGCGCCCAAGCGCGCG 17563

Result No.	Score	Query Match	Length	DB	ID	Description
1	1026	100.0	1026	5	AAD04566	Aad04566 Thermus r
2	1026	100.0	5849	5	AAD04568	Aad04568 Thermus p
3	56.8	5.5	2000	7	ADA71938	Ada71938 Rice gene
4	56.4	5.5	2030	7	ADA70342	Ada70342 Rice gene
5	56.4	5.5	2031	9	ADC08176	Adc08176 Rice DNA
6	55.4	5.4	2028	8	ADA48615	Ada48615 Rice gene
7	55.4	5.4	2028	9	ADC08212	Adc08212 Rice DNA
8	54.8	5.3	2028	7	ADA70341	Ada70341 Rice gene
9	54.8	5.3	2060	7	ADA71232	Ada71232 Rice gene
10	54.4	5.3	615	2	AAC37036	Aac37036 SOD-T' gen
11	53.2	5.2	1755	7	ADA70624	Ada70624 Rice gene
12	53.2	5.2	1965	8	ADA48223	Ada48223 Rice gene
13	53.2	5.2	1965	9	ADC08174	Adc08174 Rice DNA
14	52.4	5.1	113193	7	AAD54645	Aad54645 Streptomy
15	51.6	5.0	1815	5	AAD09400	Aad09400 Zea mays
16	50.2	4.9	509	8	ACL23248	AcL23248 DNA clone
17	50	4.9	2000	7	ADA71938	Ada71938 Rice gene
18	49.6	4.8	1674	7	ACA38419	Aca38419 Prokaryot
19	49.6	4.8	1577	7	ACB40670	AcB40670 Prokaryot
20	49.6	4.8	110000	4	AAI199682	AAI199682 27
21	49.6	4.8	110000	4	AAI199683	AAI199683 27
22	49.4	4.8	349380	6	ABQ81846	AbQ81846 Bifidobac
23	48.8	4.8	1236	9	ADC36241	AdC36241 Weed cont

CC needed for thermophilic plasmid replication. The invention relates to
CC Thermus sp. replication protein RepT, partition protein ParA and their
CC corresponding DNA molecules which relates to recombinant DNA molecules
CC encoding plasmid DNA replication origins in Thermus, as well as to
CC shuttle vectors which contain the same. The invention also relates to
CC method useful for cloning Thermus sp. plasmid genes which comprises
CC inserting plasmid DNA comprising a Thermus sp. origin of replication
CC (Ori) into a recombinant plasmid comprising a thermostable kanamycin-
CC resistance gene and an Escherichia coli Ori, to produce a cloned
CC recombinant plasmid. This cloned recombinant plasmid is transformed with
CC an E. coli. host cell, and E. coli. host cell cultured for the expression
CC of cloned recombinant plasmid. The cloned recombinant plasmid is transformed with
CC a Thermus sp. host cell is then transformed with Thermus sp. host cell and
CC from E. coli host cell is then transformed with Thermus sp. host cell and
CC Thermus sp. host cell is cultured. Thus Thermus sp. plasmid genes are
CC cloned. These plasmid DNAs are used for thermophilic transformation
XX
SQ Sequence 1026 BP; 221 A; 323 C; 285 G; 197 T; 0 U; 0 Other;
Query Match 100.0%; Score 1026; DB 5; Length 1026;
Best Local Similarity 100.0%; Pred. No. 5e-231;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGAAGACGAAAAAACCCTTCTTTAGAGAGCTTTACGAGGCTTTAGAGAAACCCACAC 60
Db 1 GTGAAGACGAAAAAACCCTTCTTTAGAGAGCTTTACGAGGCTTTAGAGAAACCCACAC 60
QY 61 AACACCGATGCCACTAGGGGGTCCAGATAGGGGGTCCAGAGGACTTCTTTGGCCACCGAC 120
Db 61 AACACCGATGCCACTAGGGGGTCCAGATAGGGGGTCCAGAGGACTTCTTTGGCCACCGAC 120
QY 121 CCCCTCCAGATGAGGTCCGAAATCCCTTCGCGAAGGCTTTACATACCAAAAAGAG 180
Db 121 CCCCTCCAGATGAGGTCCGAAATCCCTTCGCGAAGGCTTTACATACCAAAAAGAG 180
QY 181 GCATTTAGATTGCTTTACCGAGAGACCATGAGGCTTCTTCTCTCTCTGGGCC 240
Db 181 GCATTTAGATTGCTTTACCGAGAGACCATGAGGCTTCTTCTCTCTCTGGGCC 240
QY 241 CCCCTTATACACAGCTGAACCCCGTGGGAATGTATGTCAAGCCGTCCAGAACGGG 300
Db 241 CCCCTTATACACAGCTGAACCCCGTGGGAATGTATGTCAAGCCGTCCAGAACGGG 300
QY 301 CCTCAGAGTCTTGAACTCTCCAGGAGATGCCCGTCCACCATCCCTACCGCAAC 360
Db 301 CCTCAGAGTCTTGAACTCTCCAGGAGATGCCCGTCCACCATCCCTACCGCAAC 360
QY 361 CGGAGCTCTGAGAGAGGTGGGAGCGTCTTCTCATGTCCTCCCTGGAGATGTTGCC 420
Db 361 CGGAGCTCTGAGAGAGGTGGGAGCGTCTTCTCATGTCCTCCCTGGAGATGTTGCC 420
QY 421 CTCACCTGGGGTCTACCGCGAGACCGTCCAGCTCGCTGGAAGAGTCTTGAAGAAAG 480
Db 421 CTCACCTGGGGTCTACCGCGAGACCGTCCAGCTCGCTGGAAGAGTCTTGAAGAAAG 480
QY 481 GGCCTGTGCGCACCGAGCTCTTCAACAAACCGTCAACGGGAGCGCGGCGCCATCGGC 540
Db 481 GGCCTGTGCGCACCGAGCTCTTCAACAAACCGTCAACGGGAGCGCGGCGCCATCGGC 540
QY 541 ACCCTTTGGGCGTCCGGCTGAGGCGGAGGAAAGCCAGGCTCACCTGGAACGACTACATC 600
Db 541 ACCCTTTGGGCGTCCGGCTGAGGCGGAGGAAAGCCAGGCTCACCTGGAACGACTACATC 600
QY 601 TACCCCTGGAGAACTCCCTCTAGACATGGCCACCGGAGTCTCTCTTCACTGGGTC 660
Db 601 TACCCCTGGAGAACTCCCTCTAGACATGGCCACCGGAGTCTCTCTTCACTGGGTC 660
QY 661 AAGGCTTACGAGACCAACGAAATCCCGCCACACCTGGAGCTGTGCTCTCTGGCTCAG 720
Db 661 AAGGCTTACGAGACCAACGAAATCCCGCCACACCTGGAGCTGTGCTCTCTGGCTCAG 720
QY 721 GGGAAAGGCTGTATGCCACACACAGCGTGGCGGTTGACTGGGCTCTATCTGTC 780
Db 721 GGGAAAGGCTGTATGCCACACACAGCGTGGCGGTTGACTGGGCTCTATCTGTC 780

QY 781 CTCCTCCAGGTGAGCGTTTCCAACTCCCGGCCCTTATCACCTCATTTGCTACGTACATT 840
Db 781 CTCCTCCAGGTGAGCGTTTCCAACTCCCGGCCCTTATCACCTCATTTGCTACGTACATT 840
QY 841 GCCGATCTCTAGATGACCGTCTGTTCAAGACGTTTCTATGCGAGGCTTGTGTGGGCTGTG 900
Db 841 GCCGATCTCTAGATGACCGTCTGTTCAAGACGTTTCTATGCGAGGCTTGTGTGGGCTGTG 900
QY 901 GCCAGGGGTGAATCTCCCGGCAATATCTATTTCGGTCTCTAATGGGGTTATCCGAGAT 960
Db 901 GCCAGGGGTGAATCTCCCGGCAATATCTATTTCGGTCTCTAATGGGGTTATCCGAGAT 960
QY 961 TACACCGATGGCCATCTGACACCGGAGGCGTACCTAGTGAAGACCTTCAAGGAGGCC 1020
Db 961 TACACCGATGGCCATCTGACACCGGAGGCGTACCTAGTGAAGACCTTCAAGGAGGCC 1020
QY 1021 TCCTGA 1026
Db 1021 TCCTGA 1026
RESULT 2
AAD04668/c
ID AAD04668 standard; DNA; 5849 BP.
XX
AC AAD04668;
XX
DT 04-JUL-2001 (first entry)
XX
DE Thermus plasmid pTsp45S DNA sequence.
XX
KW Replication protein; RepT; partition protein; ParA; pTsp45S plasmid;
KW kanamycin-resistance gene; thermophilic transformation; Ori;
KW replication origin; ds.
XX
OS Thermus sp.
XX
FN US6207377-B1.
XX
PD 27-MAR-2001.
XX
PF 14-AUG-1998; 98US-00134246.
XX
PR 14-AUG-1998; 98US-00134246.
XX
PA (NEW) NEW ENGLAND BIOLABS INC.
XX
PI Wayne J, Xu S;
XX
DR WPI; 2001-298939/31.
XX
PT Cloning Thermus species (Ts) plasmid genes comprises transforming
PT Escherichia coli with cloned recombinant plasmid containing Ts and E.coli
PT origins of replication, isolating cloned recombinant plasmid from E.coli
PT and transforming Ts cell.
XX
PS Example 1; Fig 3; 32pp; English.
XX
CC The present sequence is Thermus plasmid pTsp45S DNA. The open reading
CC frame of pTsp45S plasmid is the replication protein. RepT which is needed
CC for thermophilic plasmid replication. The invention relates to Thermus
CC sp. replication protein RepT, partition protein ParA and their
CC corresponding DNA molecules which relates to recombinant DNA molecules
CC encoding plasmid DNA replication origins in Thermus, as well as to
CC shuttle vectors which contain the same. The invention also relates to
CC method useful for cloning Thermus sp. plasmid genes which comprises
CC inserting plasmid DNA comprising a Thermus sp. origin of replication
CC (Ori) into a recombinant plasmid comprising a thermostable kanamycin-
CC resistance gene and an Escherichia coli Ori, to produce a cloned
CC recombinant plasmid. This cloned recombinant plasmid is transformed with
CC an E. coli. host cell, and E. coli. host cell cultured for the expression
CC of cloned recombinant plasmid. The cloned recombinant plasmid is isolated

CC from E. coli host cell is then transformed with Thermus sp. host cell and
CC Thermus sp. host cell is cultured. Thus Thermus sp. plasmid genes are
CC cloned. These plasmid DNAs are used for thermophilic transformation
SQ

Sequence 5849 BP; 1377 A; 1501 C; 1670 G; 1301 T; 0 U; 0 Other;
Query Match 100.0%; Score 1026; DB 5; Length 5849;
Best Local Similarity 100.0%; Pred. No. 7.7e-231;
Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTGAAGAACGAAAAAACCCTTTTGAAGAGCTTTACAGAGCTTTAGAGGAAACCCACGAC 60
Db 4157 GTGAAGAACGAAAAAACCCTTTTGAAGAGCTTTACAGAGCTTTAGAGGAAACCCACGAC 4098

QY 61 AACACCGATGCCACTAGGGGTGAGATAGGGGTGAGAGGCTTCTTCTGGCCACCGAC 120
Db 4097 AACACCGATGCCACTAGGGGTGAGATAGGGGTGAGAGGCTTCTTCTGGCCACCGAC 4038

QY 121 CCCCTCCAGATGAGGTGCCAAATCGCTCGAAGGCTTTACATACCAAAAGAG 180
Db 4037 CCCCTCCAGATGAGGTGCCAAATCGCTCGAAGGCTTTACATACCAAAAGAG 3978

QY 181 GCACCTAGGATGCTTTACCGGAGAAAGACATAGAGCTTCTTCTCTGTGGGCC 240
Db 3977 GCACCTAGGATGCTTTACCGGAGAAAGACATAGAGCTTCTTCTCTGTGGGCC 3918

QY 241 CCCCTATACACAGCTGAAACCCCGTGGGAATGTATGTCAGAGCGTCCAGGACGGG 300
Db 3917 CCCCTATACACAGCTGAAACCCCGTGGGAATGTATGTCAGAGCGTCCAGGACGGG 3858

QY 301 CCTCAGAGCTTCTGGAATCTCTCAGAGATTCGCGCTCCACCATCCCTCAGGCAAC 360
Db 3857 CCTCAGAGCTTCTGGAATCTCTCAGAGATTCGCGCTCCACCATCCCTCAGGCAAC 3798

QY 361 CGGAGCTCTGGAGAGGTGGGACGCTGCTCTTCAATGTCCTCCCTGGAGATGTGGCC 420
Db 3797 CGGAGCTCTGGAGAGGTGGGACGCTGCTCTTCAATGTCCTCCCTGGAGATGTGGCC 3738

QY 421 CTCACCTGGGGGTCAACCGGACAGACGCTCCAGCCCTGGAAGAGGTCTTGAAGAAAG 480
Db 3737 CTCACCTGGGGGTCAACCGGACAGACGCTCCAGCCCTGGAAGAGGTCTTGAAGAAAG 3678

QY 481 GGCCTGTGGCCACGAGCTCTTCAACAAACGTCACGGGAGCGCGGCGCATCGGC 540
Db 3677 GGCCTGTGGCCACGAGCTCTTCAACAAACGTCACGGGAGCGCGGCGCATCGGC 3618

QY 541 ACCCTTTGGGCGTCCGGCTGAGGACGAAAGCCAGGCTCACCCCTGGACGACTACATC 600
Db 3617 ACCCTTTGGGCGTCCGGCTGAGGACGAAAGCCAGGCTCACCCCTGGACGACTACATC 3558

QY 601 TACCCCTGGAGAACCTCGCCCTAGACATGGCCAAACGGCGTGTCTCTCTCACTGGGTC 660
Db 3557 TACCCCTGGAGAACCTCGCCCTAGACATGGCCAAACGGCGTGTCTCTCTCACTGGGTC 3498

QY 661 AAGCCTACAGGACCAAGGATCGCCCAACCTGGACGCTGTGTGTCTGTGGCTCAG 720
Db 3497 AAGCCTACAGGACCAAGGATCGCCCAACCTGGACGCTGTGTGTCTGTGGCTCAG 3438

QY 721 GGGAAAAGGGTGATGCCCAACCAAGACCGTGGCGGTGACCTGGGCTCATCTGGTC 780
Db 3437 GGGAAAAGGGTGATGCCCAACCAAGACCGTGGCGGTGACCTGGGCTCATCTGGTC 3378

QY 781 CTCCTCGAGGTGAGCGTTCAACTCCGCGCTTATCACCCTCATCTGCTACGTACATT 840
Db 3377 CTCCTCGAGGTGAGCGTTCAACTCCGCGCTTATCACCCTCATCTGCTACGTACATT 3318

QY 841 GCGGATCTCTTAGATGACCGTGGTTCAAGAGCTTTCTATGAGCGCTTGTGTGGCTGTG 900
Db 3317 GCGGATCTCTTAGATGACCGTGGTTCAAGAGCTTTCTATGAGCGCTTGTGTGGCTGTG 3258

QY 901 GCGAGGGGTGAATCTCCCGCGCAATATCTATTTGCGCTCTAATGCGGGTTATCCGAGAT 960
Db 3257 GCGAGGGGTGAATCTCCCGCGCAATATCTATTTGCGCTCTAATGCGGGTTATCCGAGAT 3198

QY 961 TACAGGATGGCCATCTGACAGCCGGAGCGTACTAGTGAAGACCTCAAGGAGGCC 1020
Db 3197 TACAGGATGGCCATCTGACAGCCGGAGCGTACTAGTGAAGACCTCAAGGAGGCC 3138

QY 1021 TCCTGA 1026
Db 3137 TCCTGA 3132

RESULT 3
ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.
XX
AC ADA71938;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 5263.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
XX gene; ds.
XX
OS Oryza sativa.
XX
PN WO2003000898-A1.
XX
PD 03-JAN-2003.
XX
PF 22-JUN-2001; 2001WO-IB001105.
XX
PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX WPI; 2003-175290/17.
DR
XX
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 27; SEQ ID NO 5263; 899pp; English.
XX
SS The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to that
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;
Query Match 5.5%; Score 56.8; DB 7; Length 2000;
Best Local Similarity 9.1%; Pred. No. 0.0019;
Matches 60; Conservative 306; Mismatches 293; Indels 1; Gaps 1;

QY 170 ACCAAAAAGAGGACTTAGGATGCTTTACCGAAGAGACCATGAGGCTTCTCTTCT 229
Db 679 RMTARMSKRRKRWAGASMKSCWYWRGAFSWMYSKYSKACCKXKTRYMTSYNMGY 620

QY 230 CTGTGGGCCCCCTATACCACAGGTGAACCCCGCTGGGAATGTATGTCA-AGCC 288
Db 619 MYSSYSKMSWTSKMSYNGKMTCTMTSMKSTERSKXGWSKMSRMYKWKMKRKY 560

QY 289 GTCCAGGACGGGCTCAGAAAGCTTCTGGAACTCTCCAGGAGATTGCCCGCTCCACCATC 348

```

Db 559 RYMKWCKTWRRCWCVWGYTMTTSSRMVYGRYKARYTSRRYMYKRYKYWYVY 500
QY 349 CCTACGGCAACGGGAGCTCTGGAGGAGTGGGAGGTGCTCTTATGTCCTCCCTG 408
Db 499 MYMKCSYMRVYGCACCKCCCTACWKAAYSMMYVYRKYKWMRSTKYWMSWYK 440
QY 409 GAGATTTGGCCCTCAACCTGGGGTCAACCGGCAAGCGTCCACGCTGGAAGATC 468
Db 439 RSMKYGAKGCGCKWYVYCSYVYKMYVYVYKMYVYVYKMYVYVYKMYVYVY 380
QY 469 CTGAGAAAAGGGCTGTGTCGCCACCGAGCTCTTCAACCAACCGTCAACGGGAGCG 528
Db 379 SMWTWYVYVYKMYVYKRGVYVYKMYVYKRYKMYVYKMYVYKMYVYKMYVY 320
QY 529 CGGGCCATCGGACCCCTTTGGGCGCTTCCGGCTGAGCGGAGGAAAGCCAGGCTCACCTG 588
Db 319 KRCYCHRYATCYWCCYKRGVYVYKMYVYKRYKMYVYKMYVYKMYVYKMYVY 260
QY 589 GAGCACTATACCTACCCCTGGAGGAACTCGCGCTAGACATGCGCAAGCGCTGCTCC 648
Db 259 YVMSYVWARSSTWRSAAKRYKGYSTSRRAKMYVYKMYVYKMYVYKMYVYK 200
QY 649 TTCACTGGTCAAGGCTTACGAGGACCAAGCAATCCCGCCACCGCTGAGCTGCTGTC 708
Db 199 SSKWYKNSKSKMYVYKSSWCCCTCYVYKMYVYKMYVYKMYVYKMYVYKMYVY 140
QY 709 CTCTGGCTCAGGGGAAAGGTGATGTCACCAACCAAGACCGTGGCGGTGACCTGGC 768
Db 139 MCKKYCSCTKYCSYVYKMYVYKMYVYKMYVYKMYVYKMYVYKMYVYKMYVY 80
QY 769 CTATCTCTGCTCTCCCGGAGTGGAGCTTCCAACTCCCGGCTTATCACCTCAT 828
Db 79 MSCAYSTSTSRWMSMYVYKMYVYKMYVYKMYVYKMYVYKMYVYKMYVYKMYVY 20

```

RESULT 4

```

ID ADA70342 standard; DNA; 2030 BP.
XX ADA70342;
XX AC ADA70342;
XX AC ADA70342;
DT 20-NOV-2003 (first entry)
XX DE Rice gene, SEQ ID 3665.
XX KW Plant; bacterial infection; fungal infection; viral infection; rice;
XX KW gene; ds.
XX OS Oryza sativa.
XX PN WO2003000898-A1.
XX PD 03-JAN-2003.
XX PF 22-JUN-2001; 2001WO-IB001105.
XX PR 22-JUN-2001; 2001WO-IB001105.
XX PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
XX PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX DR WPI; 2003-175290/17.
XX PT Identifying at least one gene involved in plant resistance or response to
XX PT pathogenic infection for conferring resistance or tolerance to a plant to
XX PT bacterial, fungal or viral infection by determining or detecting plant
XX PT gene expression.
XX PS Claim 6; SEQ ID NO 3665; 899pp; English.
XX PF

```

```

CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2030 BP; 388 A; 659 C; 727 G; 246 T; 0 U; 10 Other;

Query Match 5.5%; Score 56.4; DB 7; Length 2030;
Best Local Similarity 46.7%; Pred. No. 0.0023;
Matches 214; Conservative 0; Mismatches 241; Indels 3; Gaps 1;

QY 356 GCAACCGGGAGCTCTGGAGGAGGTGGGACGGTCTTTCATGTGTCCTCCCTGGAGATGT 415
Db 485 GCGCCATGGTGTCTACGCGGATGAGGAGAGCGGCCGAGGCTACCTCGCGGAGAGGTCA 544
QY 416 TGGCCCTCAACCTGGGGGTCAACCGGACAGCTCCACGCTGGAAGAGGTCTTGA 475
Db 545 CGCGCGCGCTGCTCACCGTCCCGGCTTCTTCAACGACGCGGAGGCGGACCAAG 604
QY 476 AAAAGGCGCTGGTGGCCACCGAGCTCTTCAACCAACCGTCAACGGGGAGCGCGGCCCA 535
Db 605 AGCGCGGCTCATCGCGGGCTCACGTCGACGCGCATCATCAACGAGCGGACCGCGCG 664
QY 536 TCGGCAACCTTTGGGCGCTCGGCTGAGGCGAGGAAAGCGAGCTCACCTGGAGCAT 595
Db 665 CCATCGCTACGCGATCGACAGAGGCGCGGAGAGAACGCTCTCTTTCGAGCTCG 724
QY 596 ACATCTACCTCTGGAGAACCTCGCCTAGACATGCAACGCGGCTCTCTCTCAACT 855
Db 725 CGCGGCGCATGTCGAGCTAGCATCTCTGCCATCGACAGCGGCTTCGAGGTCCTTG 784
QY 656 GGGTCAAGGCTTACCGAGGACCGGAAATC---CGCCCCACCTGGAGCTGCTCTCT 712
Db 785 CCACCAACGGCGACACCCACCTCGGCGGAGGACTTCGACCAACGCTCATGGAGCACT 844
QY 713 GGGCTAGGGGAAAGGCTGATGCCCAACACGAGACCGTGGCGCTTGACCTGGGCTCA 772
Db 845 TCGTCAAGGTATCCCGCGGAGACGCGGCGGACATCGCGCGGAGCGCGGCGCTGG 904
QY 773 TCGTGGTCTCCCGGAGGTGGAGCTTCCAAACTCCCG 810
Db 905 GCAAGCTCCCGGAGTGGAGCGCGCAAGCGCGG 942

RESULT 5
ADCO8176
ID ADC08176 standard; DNA; 2031 BP.
XX ADC08176;
XX AC ADC08176;
XX AC ADC08176;
DT 18-DEC-2003 (first entry)
XX DE Rice DNA sequence Seq ID443 related to grain filling.
XX PA plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
XX KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
XX KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
XX KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
XX KW gene; ds; plant.
XX OS Oryza sativa.
XX PN WO2003000905-A2.
XX PD 03-JAN-2003.
XX PS 21-JUN-2002; 2002WO-IB002450.
XX PF

```

XX 22-JUN-2001; 2001US-0300112P.
PR 26-SEP-2001; 2001US-0325277P.
PR 20-DEC-2001; 2001US-0342327P.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Glazebrook J, Katagiri F, Kreps J, Provart N, Rické D;
XX
XX WPI; 2003-229341/22.
DR P-PSDB; ADC08177.
XX
XX New plant genes encoding polypeptides having an activity involved in or
PT associated with the synthesis, metabolism or degradation of carbohydrates
PT in the plant grain useful in generating plants having improved
PT nutritional properties.
XX
XX Claim 35; SEQ ID NO 443; 130pp; English.
XX
XX This invention, in the area of plant biotechnology, relates to novel
CC polynucleotides comprising a nucleotide sequence encoding a protein which
CC is involved in or associated with the synthesis, metabolism or
CC degradation of carbohydrates in the plant grain and the expression of
CC which is up-regulated during grain filling. The plant is selected from
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC sugarbeet, wheat, and rice. The invention may be useful for the
CC improvement of protein, oil, starch, fibre and moisture content of the
CC cereal grains. In addition, carbohydrate levels may be modified to a more
CC desirable level using the present invention. The present sequence is a
CC DNA sequence encoding a rice protein of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/publishedpct_sequences.
XX
XX Sequence 2031 BP; 388 A; 659 C; 728 G; 246 T; 0 U; 10 Other;
SQ

Query Match 5.4%; Score 56.4; DB 9; Length 2031;
Best Local Similarity 46.7%; Pred. No. 0.0023;
Matches 214; Conservative 0; Mismatches 241; Indels 3; Gaps 1;
XX
QY 356 GCAACCGGAGCTCTGGAGGAGTGGGAGCGTCTTTCATGGTCCCTGGAGATGT 415
Db 485 GCGCATGTGTCTACCGGATGAAGAGACGCGCGGCTACCTCTGGGAGAGTCA 544
QY 416 TGGCCCTTAACCTGGGGGTCAACCGGACGCTCCACGCTTGAAGAGTCTTTGAGA 475
Db 545 CGCGGCGCTCTACCGTCCCGCTTACTTCAACGACGCGCAGGCCACCAAGG 604
QY 476 AARAGGCGTGGTGGCCACCGACGCTCTTCAACAAACGTCACGGGAGCGCGGCCA 535
Db 605 ACGCGGCGTCACTCGCGGGTCAACCGTCAACCGATCATCAACGACCGACCGCGCG 664
QY 536 TCGGCACCTTTGGGCGCTCGGGTGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 595
Db 665 CCATCGCTACGGCATCGACAAGAGGCGCGGAGAGAGGAGGAGGAGGAGGAGGAG 724
QY 596 ACATCTACCTCTGAGGAACTCCGCTTACATAGACATGGGCGAAGCGGCTCTCTCACT 655
Db 725 GCGGCGGACGTTTCGAGTCTCGACATCTCTCGCATTCGACACGCGGCTTTCGAGTCTTG 784
QY 656 GGGTCAAGGCTTACAGACCAACGGAATC---CGCCCCACCTGGAGCTGTGTCTCT 712
Db 785 CCACCAACGGGACACCCACCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 844
QY 713 GGGTCAAGGAAAGGATGATGCCCAACCAAGACCGTGGCGGTGAGCTGGGCTCA 772
Db 845 TCGTCAAGGTATCTCGCGGAAAGACGCGGCGGACATCGCGGCGAGCGCGCGGTGG 904
QY 773 TCCTGGTCTCTCCCGAGTGGAGCGTTCCAACTCCG 810
Db 905 GCAAGCTCCGCGGAGTGGAGCGCGCCCAAGCGCGG 942

RESULT 6
ADA48615
ID ADA48615 standard; DNA; 2028 BP.
XX
XX ADA48615;
XX
XX 20-NOV-2003 (first entry)
XX
XX Rice gene conferring disease resistance in plants.
XX
XX disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant.
XX
XX Oryza sativa.
XX
XX WO200300906-A2.
XX
XX 03-JAN-2003.
XX
XX 21-JUN-2002; 2002WO-1B002453.
XX
XX 22-JUN-2001; 2001US-0300112P.
XX
XX 28-SEP-2001; 2001US-0352277P.
XX
XX 22-MAR-2002; 2002US-0366535P.
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX
XX Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Katagiri F, Kreps J, Provart N, Rické D, Zhu T;
XX
XX WPI; 2003-184052/18.
DR P-PSDB; ADA48616.
XX
XX New polynucleotide comprising a plant nucleotide sequence having an open
PT reading frame that encodes a polypeptide associated with disease
PT resistance, useful for conferring resistance or tolerance to a plant
PT pathogen.
XX
XX Claim 1; SEQ ID NO 685; 299pp; English.
XX
XX The invention relates to a novel isolated polynucleotide comprising a
CC plant nucleotide sequence having an open reading frame that encodes a
CC polypeptide associated with disease resistance or its fragment having
CC substantially the same activity as the full-length polypeptide. The
CC polynucleotide of the invention is useful for conferring resistance or
CC tolerance to a plant pathogen. The present sequence represents a gene
CC conferring disease resistance used in the invention.
XX
XX Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;
SQ

Query Match 5.4%; Score 55.4; DB 8; Length 2028;
Best Local Similarity 46.5%; Pred. No. 0.004;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;
QY 356 GCAACCGGAGCTCTGGAGGAGTGGGAGCGTCTTTCATGGTCCCTGGAGATGT 415
Db 485 GCGCATGTGTCTACCGGATGAAGAGACGCGCGGCTACCTCTGGGAGAGTCA 544
QY 416 TGGCCCTTAACCTGGGGGTCAACCGGACGCTCCACGCTTGAAGAGTCTTTGAGA 475
Db 545 CGCGGCGCTCTACCGTCCCGCTTACTTCAACGACGCGCAGGCCACCAAGG 604
QY 476 AARAGGCGTGGTGGCCACCGACGCTCTTCAACAAACGTCACGGGAGCGCGGCCA 535
Db 605 ACGCGGCGTCACTCGCGGGTCAACCGTCAACCGATCATCAACGACCGACCGCGCG 664
QY 536 TCGGCACCTTTGGGCGCTCGGGTGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 595
Db 665 CCATCGCTACGGCATCGACAAGAGGCGCGGAGAGGAGGAGGAGGAGGAGGAGGAG 724
QY 596 ACATCTACCTCTGAGGAACTCTCGGCTTAGACATGGGCGAAGCGGCTCTCTCACT 655
Db 725 GCGGCGGACGTTTCGAGTCTCGACATCTCTCGCATTCGACACGCGGCTTTCGAGTCTTG 784
QY 773 TCCTGGTCTCTCCCGAGTGGAGCGTTCCAACTCCG 810
Db 905 GCAAGCTCCGCGGAGTGGAGCGCGCCCAAGCGCGG 942

QY 656 GGGTCAAGGCTTACGAGGACACGGAATC---CGCCCCACCTGGACGTGCTGCTCTCT 712
 DB 785 CCACCAACGGCGACACCCACCTCGCGCGGAGGACTTTCGACCAACGCTCATGACCACT 844
 QY 713 GGGCTCAGGGAAGGATGATGCCCAACACACAGACCGTGGCGGTGACCTGGGCTCA 772
 DB 845 TCGTCAAGGTATCGCGCGGAAGCACGGGCGGACATCGCGCGGACGCGCGCGCTGG 904
 QY 773 TCCTGTCTCCCGAGGTGAGGCTTCCAACTCCCG 810
 DB 905 GCAAGCTCGCGCGAGTGGAGCGCGCAAGCGCGG 942

RESULT 7

ADC08212

ID ADC08212 standard; DNA; 2028 BP.

XX AC

XX ADC08212;

XX AC

XX 18-DEC-2003 (first entry)

XX DT

XX DE

XX Rice DNA sequence Seq ID517 related to grain filling.

XX KW

XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;

XX KW

XX carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;

XX KW

XX tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;

XX KW

XX wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;

XX KW

XX gene; ds; plant.

XX OS

XX Oryza sativa.

XX PN

XX WO2003000905-A2.

XX XX

XX 03-JAN-2003.

XX XX

XX 21-JUN-2002; 2002WO-IB002450.

XX XX

XX 22-JUN-2001; 2001US-0300112P.

XX PR

XX 26-SEP-2001; 2001US-0325277P.

XX PR

XX 20-DEC-2001; 2001US-0342327P.

XX XX

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PA

XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;

XX PI

XX Glazebrook J, Katagiri F, Kreps J, Provart N, Rickes D;

XX XX

XX P-PSDB; ADC08213.

XX DR

XX WPI; 2003-229341/22.

XX DR

XX Claim 35; SEQ ID NO 517; 130pp; English.

XX FS

XX XX

XX This invention, in the area of plant biotechnology, relates to novel

XX CC

XX polynucleotides comprising a nucleotide sequence encoding a protein which

XX CC

XX is involved in or associated with the synthesis, metabolism or

XX CC

XX degradation of carbohydrates in the plant grain and the expression of

XX CC

XX which is up-regulated during grain filling. The plant is selected from

XX CC

XX corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,

XX CC

XX sugarbeet, wheat, and rice. The invention may be useful for the

XX CC

XX improvement of protein, oil, starch, fibre and moisture content of the

XX CC

XX cereal grains. In addition, carbohydrate levels may be modified to a more

XX CC

XX desirable level using the present invention. The present sequence is a

XX CC

XX DNA sequence encoding a rice protein of the invention. Note: The sequence

XX CC

XX data for this patent did not form part of the printed specification, but

XX CC

XX was obtained in electronic format directly from WIPO at

XX CC

XX ftp.wipo.int/pub/publishedpct_sequences.

XX CC

XX Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;

XX SQ

Query Match 5.4%; Score 55.4; DB 9; Length 2028;
 Best Local Similarity 46.5%; Pred. No. 0.004;
 Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;
 QY 356 GCACCGGGAGCTCTGGAGAGGTGGGACGGTCTTCTATGTGTCCTCCCTGGAGATGT 415
 DB 485 GCGCCATGTGTCTACCCCGGATGAAGAGAGACGGCCCTTACCTCGCGGAGAGGTCA 544
 QY 416 TGGCCCTCAACCTGGGGGTACCCCGGACAGCCGTCACGCTTGGAGAGGTCTTGAGA 475
 DB 545 CGCGGCGCGTCTGTCACCGTCCCGGCTTCTTCAACGACGCGGAGCGCCACCAAG 604
 QY 476 AAAAGGGCTGTGGCCACCGAGCTCTTTCACCAACCGTCAACGGGAGCGCGGCCA 535
 DB 605 ACGCCGCGGTCTATCGCGCGGGCTCACCGTCAACCGATCATCAACGAGCGCGCG 664
 QY 536 TCGGCACCTTTTGGGCGCTCCGCTGAGCCAGGAGAAAGCCAGGCTCACCTTGGACGACT 595
 DB 565 CCATCGCTACGCTGACAGAGAGGGCGGAGAGAAAGCTCTCTCTTGCACCTCG 724
 QY 596 ACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCAAAGCGCTCTCTCTTCAACT 655
 DB 725 GCGCGCGGACGTTTCGACGTGACATCTCGCCATCGACAAACGCGTGTTCGAGGTCTTG 784
 QY 656 GGGTCAAGGCTTACGAGGACCGGAAATC---CGCCCAACCTGGAGCTGCTGTCCTCT 712
 DB 785 CCACCAAGCGGACACCCACTTGGCGGAGGACTTTCGACCAAGCGCTCATGAGCACT 844
 QY 713 GGGCTCAGGGAAGGGTGTATGCCCAACACCAAGCGTGGCGCTTCACTGGGCTCA 772
 DB 845 TCGTCAAGGTCTCCCGGAGGTGGAGCGTTCCAACTCCCG 810
 QY 773 TCCTGTCTCTCCCGGAGGTGGAGCGTTCCAACTCCCG 810
 DB 905 GCAAGCTCGCGCGGAGTTCGAGCGCGCCGACCGCGCGCTGG 942

RESULT 8

ADA70341

ID ADA70341 standard; DNA; 2028 BP.

XX AC

XX ADA70341;

XX AC

XX ADA70341;

XX DT

XX 20-NOV-2003 (first entry)

XX XX

XX DE

XX Rice gene, SEQ ID 3664.

XX XX

XX KW

XX Plant; Bacterial infection; fungal infection; viral infection; rice;

XX KW

XX gene; ds.

XX XX

XX OS

XX Oryza sativa.

XX XX

XX PN

XX WO2003000898-A1.

XX XX

XX 03-JAN-2003.

XX XX

XX 22-JUN-2001; 2001WO-IB001105.

XX PF

XX 22-JUN-2001; 2001WO-IB001105.

XX PR

XX (SYGN) SYNGENTA PARTICIPATIONS AG.

XX PA

XX Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

XX PI

XX Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

XX DR

XX WPI; 2003-175290/17.

XX XX

XX Identifying at least one gene involved in plant resistance or response to

XX PT

XX pathogenic infection for conferring resistance or tolerance to a plant to

XX PT

XX bacterial, fungal or viral infection by determining or detecting plant

XX PT

XX gene expression.

XX XX

```
PS Claim 6; SEQ ID NO 3664; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2028 BP; 390 A; 662 C; 729 G; 246 T; 0 U; 1 Other;

Query Match 5.3%; Score 54.8; DB 7; Length 2028;
Best Local Similarity 46.5%; Pred. No. 0.0056;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

QY 356 GCAACCGGGAGCTCTGGAGGAAGTGGGGACGGTCTTTCATGTCCTCCCTGGAGATGT 415
DB 485 GCGCATGTGCTACCCGGATGAAGAGACGGCGAGGCTACTCTGGCGAGAGGTCA 544
QY 416 TGGCCCTCAACTGGGGGTACCGCGAGACCGTCCAGCGCTGGAAGAGTCTTTGAGA 475
DB 545 CGCGCGCGTCTGTCACCGTCCCGGCTACTTCAACGACGCGCAGCGGACCAAGG 604
QY 476 AAAGGGGCTGTGGCCACCGAGCTCTTACCAAAACCGTCAACGGGAGCGCGGCCCA 535
DB 605 ACGCGGGCTCATCGCGGGTCACTCGTCCGATCATCAACGACGCGACGCCCGCG 664
QY 536 TCGGACCCCTTTGGCCGTCCGGTGAAGCCAGGAAAGCCAGGCTACCCCTGGACACT 595
DB 665 CCATCGCTACGGCATCGACAAGAGGGCGCGAGAGAACTCTCTGCTTCGACCTCG 724
QY 596 ACATCTACCCCTGGAGAACCTCGCCCTAGACATGCGCAAGCGGCTCTCTTCACT 655
DB 725 GCGCGGCACGTTCGACGTACGATCTTCGCTGATCGACACGCGGTGTTCAGGTCTTG 784
QY 656 GGGTCAAGGCTTACAGGACCAAGATCCG---CCCCACCCCTGGAGTGTGTCCTCT 712
DB 785 CCACCAACGGCGACACCCACTCGGCGGAGGACTTCGACCAACGCTCATGACCACT 844
QY 713 GGGTCAAGGAAAGGTGATGCCACACAGACGCTGGCGCTTGACCTGGGCTCA 772
DB 845 TCGTCAAGGTTCATCGCGGAGGACGCGGCGGACATCGCGGCGACGCGCGCGCTGG 904
QY 773 TCCTGTCCTCCCGAGTGGAGGTTCCAAACTCCCG 810
DB 905 GCAAGCTCCGCGAGTGGAGGCGGCGCAAGCGCGG 942

RESULT 9
ADA71232
ID ADA71232 standard; DNA; 2060 BP.
XX
AC ADA71232;
XX
DT 20-NOV-2003 (first entry)
XX
DE Rice gene, SEQ ID 4555.
XX
KW Plant; bacterial infection; fungal infection; viral infection; rice;
KW gene; ds.
XX
OS Oryza sativa.
XX
PN WO200300898-A1.
XX
PD 03-JAN-2003.
XX
XX 22-JUN-2001; 2001WO-IB001105.
XX

PR 22-JUN-2001; 2001WO-IB001105.
XX
PA (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;
PI Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;
XX
DR WPI; 2003-175290/17.
XX
PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.
XX
PS Claim 6; SEQ ID NO 4555; 899pp; English.
XX
CC The present invention relates to a method (M1) for identifying genes
CC involved in plant resistance or response to pathogenic infection. M1
CC comprises identifying a gene whose expression is significantly altered in
CC the incompatible interaction of plant gene expression relative to
CC expression of the gene in an uninfected plant, in a mutant plant that
CC does not express a gene associated with response to pathogenic infection,
CC or in a corresponding incompatible or compatible interaction. (M1) is
CC useful for conferring resistance to resistance or tolerance to a plant to
CC bacterial, fungal or viral infection. The present sequence was used to
CC illustrate the invention.
XX
SQ Sequence 2060 BP; 395 A; 670 C; 739 G; 254 T; 0 U; 2 Other;

Query Match 5.3%; Score 54.8; DB 7; Length 2060;
Best Local Similarity 46.5%; Pred. No. 0.0056;
Matches 213; Conservative 0; Mismatches 242; Indels 3; Gaps 1;

QY 356 GCAACCGGGAGCTCTGGAGGAAGTGGGGACGGTCTTTCATGTCCTCCCTGGAGATGT 415
DB 485 GCGCATGTGCTACCCGGATGAAGAGACGGCGAGGCTACTCTGGCGAGAGGTCA 544
QY 416 TGGCCCTCAACTGGGGGTACCGCGAGACCGTCCAGCGCTGGAAGAGTCTTTGAGA 475
DB 545 CGCGCGCGTCTGTCACCGTCCCGGCTACTTCAACGACGCGCAGCGGACCAAGG 604
QY 476 AAAGGGGCTGTGGCCACCGAGCTCTTACCAAAACCGTCAACGGGAGCGCGGCCCA 535
DB 605 ACGCGGGCTCATCGCGGGTCACTCGTCCGATCATCAACGACGCGACGCCCGCG 664
QY 536 TCGGACCCCTTTGGCCGTCCGGTGAAGCCAGGAAAGCCAGGCTACCCCTGGACACT 595
DB 665 CCATCGCTACGGCATCGACAAGAGGGCGCGAGAGAACTCTCTGCTTCGACCTCG 724
QY 596 ACATCTACCCCTGGAGAACCTCGCCCTAGACATGCGCAAGCGGCTCTCTTCACT 655
DB 725 GCGCGGCACGTTCGACGTACGATCTTCGCTGATCGACACGCGGTGTTCAGGTCTTG 784
QY 656 GGGTCAAGGCTTACAGGACCAAGATCCG---CCCCACCCCTGGAGTGTGTCCTCT 712
DB 785 CCACCAACGGCGACACCCACTCGGCGGAGGACTTCGACCAACGCTCATGACCACT 844
QY 713 GGGTCAAGGAAAGGTGATGCCACACAGACGCTGGCGCTTGACCTGGGCTCA 772
DB 845 TCGTCAAGGTTCATCGCGGAGGACGCGGCGGACATCGCGGCGACGCGCGCGCTGG 904
QY 773 TCCTGTCCTCCCGAGTGGAGGTTCCAAACTCCCG 810
DB 905 GCAAGCTCCGCGAGTGGAGGCGGCGCAAGCGCGG 942

RESULT 10
AAQ37036
ID AAQ37036 standard; DNA; 615 BP.
XX
AC AAQ37036;
XX
DT 05-JUL-1993 (first entry)
```


XX ADA48223;
AC 20-NOV-2003 (first entry)
DT Rice gene conferring disease resistance in plants.
DE disease resistance; pathogen tolerance; plant pathogen; ds; gene; plant.
DS Oryza sativa.
KW WO2003000906-A2.
XX 03-JAN-2003.
XX 21-JUN-2002; 2002WO-IB002453.
XX 22-JUN-2001; 2001US-0300112P.
PR 26-SEP-2001; 2001US-0352277P.
PR 22-MAR-2002; 2002US-0366535P.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Glazebrook J, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Katagiri F, Kreps J, Provart N, Ricke D, Zhu T;
XX WPI; 2003-184052/18.
DR P-PSDB; ADA48224.
XX
XX New polynucleotide comprising a plant nucleotide sequence having an open
PT reading frame that encodes a polypeptide associated with disease
PT resistance, useful for conferring resistance or tolerance to a plant
PT pathogen.
XX
XX Claim 1; SEQ ID NO 293; 299pp; English.
XX
XX The invention relates to a novel isolated polynucleotide comprising a
CC plant nucleotide sequence having an open reading frame that encodes a
CC polypeptide associated with disease resistance or its fragment having
CC substantially the same activity as the full-length polypeptide. The
CC polynucleotide of the invention is useful for conferring resistance or
CC tolerance to a plant pathogen. The present sequence represents a protein
CC conferring disease resistance used in the invention.
XX
XX Sequence 1965 BP; 383 A; 646 C; 699 G; 235 T; 0 U; 2 Other;
SQ
Query Match 5.2%; Score 53.2; DB 8; Length 1965;
Best Local Similarity 46.3%; Pred. No. 0.013;
Matches 212; Conservative 0; Mismatches 243; Indels 3; Gaps 1;
QY 356 GCACCGGAGCTCTGGAGGAGGTGGGACGGTCTTTCATGGTCCCTCTGGAGATGT 415
DB 437 GCGCCATGGTCTCACGCGGATGAAGGAGAGCGGCGGACCTACCTCGGAGAGGTCA 496
QY 416 TGGCCCTCAACTGGGGTCAACCGGAGACCGTCCAGCTGGAAGATCTTCTGAGA 475
DB 497 CCGGCGGCTGTCACCGTCCCGCTTACTTCAACGAGCGGACGCGAGCCACCAAGG 556
QY 476 AAAAGGCGCTGGTGCCACCGACGCTCTTCAACAAACCGTCAACGGGAGCGCGGCGCA 535
DB 557 ACGCGCGGCTCATCGCGGGCTCACCGTGCACCGTGCACGATCATCAACGAGCCACCGCGCG 616
QY 536 TCGGACCTTTGGCGCTCGGCTGAGCGGAGGAGCCAGCTCACCTGGAGACT 595
DB 617 CCATCGCTACGGCATCGACAAGAGGCGCGGAGAGACGTCCTGCTTGCACCTCG 676
QY 596 ACATCTACCTTGGAGAACCTCGCCCTTAGACATGGGCAACGGCGTCTCTCTTCAACT 655
DB 677 GCGGCGGACGCTCGACGTACGATCTCTCGCCATCGACACGGGCTGTTGAGTCTCTG 736
QY 656 GGGTCAAGGCTTACAGAGCAACCGAATCG- - -CCCAACCTGGAGCTGCTCTCT 712
DB 737 CCACCAAGGCGGACCAACCACTCGGCGGCGGAGGACTTCGACCAACGCTCATGACCACT 796

QY 713 GGGCTCAGGGGAAAAGCGGTGATGCCCAACACCAAGACCGTGGCGGTGACCTGGGCTCA 772
DB 797 TCGTCAAGGTCTATCCGCGGAAAGCAGCGCGGACATCACCGCGACGCGCGCTGG 856
QY 773 TCGTGTCTCTCCCGAGGTGGAGCGTTCCAAACTCCCG 810
DB 857 GCAAGCTCCGCGGAGTCTCGAGCGCGGCAAGCGCGG 894
RESULT 13
ADC08174
ID ADC08174 standard; DNA; 1965 BP.
XX AC ADC08174;
XX 18-DEC-2003 (first entry)
DE Rice DNA sequence Seq ID441 related to grain filling.
XX plant biotechnology; carbohydrate synthesis; carbohydrate metabolism;
KW carbohydrate degradation; carbohydrate; plant grain; grain filling; corn;
KW tomato; banana; canola; cotton; peanut; sorghum; tobacco; sugarbeet;
KW wheat; rice; protein; oil; starch; fibre; moisture content; cereal grain;
KW gene; ds; plant.
XX Oryza sativa.
XX WO2003000905-A2.
XX 03-JAN-2003.
XX 21-JUN-2002; 2002WO-IB002450.
XX 22-JUN-2001; 2001US-0300112P.
PR 26-SEP-2001; 2001US-0352277P.
PR 20-DEC-2001; 2001US-0342327P.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX Zhu T, Cheng W, Briggs S, Cooper B, Goff SA, Moughamer T;
PI Glazebrook J, Katagiri F, Kreps J, Provart N, Ricke D;
XX WPI; 2003-229341/22.
DR P-PSDB; ADC08175.
XX
XX New plant genes encoding polypeptides having an activity involved in or
PT associated with the synthesis, metabolism or degradation of carbohydrates
PT in the plant grain useful in generating plants having improved
PT nutritional properties.
XX Claim 35; SEQ ID NO 441; 130pp; English.
XX
XX This invention, in the area of plant biotechnology, relates to novel
CC polynucleotides comprising a nucleotide sequence encoding a protein which
CC is involved in or associated with the synthesis, metabolism or
CC degradation of carbohydrates in the plant grain and the expression of
CC which is up-regulated during grain filling. The plant is selected from
CC corn, tomato, banana, canola, cotton, peanut, sorghum, tobacco,
CC sugarbeet, wheat, and rice. The invention may be useful for the
CC improvement of protein, oil, starch, fibre and moisture content of the
CC cereal grains. In addition, carbohydrate levels may be modified to a more
CC desirable level using the present invention. The present sequence is a
CC DNA sequence encoding a rice protein of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/publishedpct_sequences.
XX
SQ Sequence 1965 BP; 383 A; 646 C; 699 G; 235 T; 0 U; 2 Other;
Query Match 5.2%; Score 53.2; DB 9; Length 1965;
Best Local Similarity 46.3%; Pred. No. 0.013;
Matches 212; Conservative 0; Mismatches 243; Indels 3; Gaps 1;

CC responsible for the biosynthesis of the polyene antibiotic amphotericin
CC (amph) of Streptomyces nodosus. Polynucleotides of the invention are
CC useful for preparing amphotericin derivatives or analogue antibiotic
CC agents with altered properties and in the biosynthesis of polyketides
CC other than amphotericin. amphDII, amphDII or amphDI mutants are useful
CC for producing amphotericin derivatives glycosylated with alternative
CC sugars; amphDII or amphDII gene sequences are useful in engineered
CC biosynthesis of perosaminyl-amphoteronolide B; amphDII or amphDII and
CC amphN gene sequences are useful in the engineered biosynthesis of
CC perosaminyl-16-desacarbonyl-16-methyl amphoteronolide B; amphDII, amphDII
CC and amphDI gene sequences are useful for preparing polyketides capable
CC of addition of mycosamine to a polyketide other than amphoteronolide A or
CC B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
CC The present sequence is S. nodosus amph biosynthetic gene cluster
XX
SQ Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;
Query Match 5.1%; Score 52.4; DB 7; Length 113193;
Best Local Similarity 46.9%; Pred. No. 0.056;
Matches 164; Conservative 0; Mismatches 186; Indels 0; Gaps 0;
QY 320 TCCTCCAGGAGATTGCGCGTCCACCATCCCTTACGAGATGTCCTCAACCTGGGGTCAACC 379
Db 15513 TCGCCCTGACCTGCGCTGCCAGCGGTGCGCAACGCGAGAGCAACTGCGGTCCGCG 15572
QY 380 TGGGAGCGGTCTTCTCATGTCTCCCTGGAGATGTCCTCAACCTGGGGTCAACC 439
Db 15573 CGCGGCCACCGTATGACACCCCAACGCGGTTCGTGGCGTTGAGCGGCGCGCC 15632
QY 440 GCGAGACCGTCCAGCGCTGGAAGAGTCTCTTGAGAAAAGGGCTGTGGCCACCGACG 499
Db 15633 TCGCGAGGAGCGCGGTGCAAGCGTTCTCCGAGAGCGCGCATGACCTCGCG 15692
QY 500 TCCTTCACCAACCGTCAACGGGAGCGCGGCCATCGGCACCTTTGGCGCTCGGC 559
Db 15693 AAGGTGTGCGCATCGTCTCTGAGCGGTGTCGCAACGCGCGCGCAACGCGCACCGG 15752
QY 560 TGAGGCCAGGAAAGCAGGCTCACTCCCTGGACGACTACATCTACCCCTGGAGGAACCTCG 619
Db 15753 TGCTCGCGGTATCCGCGGCTCCGCCATCAACAGAGCGGCGCTCCAAAGGCTCAGG 15812
QY 620 CCTTAGACATGGCAACGGCGTGTCTCTCTTCAACTGGGTCAAGGCTAC 669
Db 15813 CCCCCAACGGCCCTCCCAAGCAGCGCGTCTATCCGCGAGCGCGCTGCCCAAC 15862
RESULT 15
AAD09400 standard; cDNA; 1815 BP.
XX
AC AAD09400;
XX
DT 10-SEP-2001 (first entry)
XX
DE Zea mays neoxanthin cleavage enzyme, VP14 cDNA.
XX
KW Maize; neoxanthin cleavage enzyme; VP14; abscisic acid; ABA; herbicide;
KW stress tolerance; transgenic plant; plant breeding; antisense-therapy;
KW plant growth protectant; ss.
XX
OS Zea mays.
XX
FH Key Location/Qualifiers
FT 1. 1815
FT /*tag= a
FT /product= "Zea mays VP14 protein"
XX
PN EP1116794-A2.
XX
PD 18-JUL-2001.
XX
PF 11-JAN-2001; 2001EP-00300218.

PR 13-JAN-2000; 2000JP-00010056.
PR 11-JAN-2001; 2001JP-00003476.
XX
PA (RIKE) RIKEN KK.
XX
PI Tuchi S, Kobayashi M, Shinozaki K;
XX WPI; 2001-400081/43.
DR P-PSDB; AAE04788.
XX
PT A DNA encoding a protein with a neoxanthin cleavage activity for
PT producing transgenic plants with improved or decreased stress tolerance.
XX
PS Claim 3; Page 60-64; 101pp; English.
XX
CC The invention relates to neoxanthin cleavage enzymes and their
CC corresponding cDNA molecules. Neoxanthin cleavage enzyme plays a key role
CC in endogenous abscisic acid (ABA) biosynthesis under drought stress.
CC Neoxanthin cleavage enzyme is used for improving stress tolerance in a
CC plant when expressed in a plant cell. The invention also relates to
CC methods for increasing or decreasing stress tolerance in a plant by
CC introducing the DNA into the plant, and a transgenic plant into which a
CC neoxanthin cleavage enzyme is introduced. The improvement of stress
CC tolerance in plants is useful, for example in plant breeding. Neoxanthin
CC cleavage enzyme genes are useful for producing transgenic plants. An arid
CC land can be improved by growing transformant weed for several years and
CC then removing the weed by specifically lowering stress tolerance in the
CC weed by inducing an inducible promoter. The present cDNA sequence encodes
CC Zea mays neoxanthin cleavage enzyme, VP14 protein related to the
CC invention
XX
SQ Sequence 1815 BP; 266 A; 689 C; 604 G; 256 T; 0 U; 0 Other;
Query Match 5.0%; Score 51.6; DB 5; Length 1815;
Best Local Similarity 45.9%; Pred. No. 0.031;
Matches 177; Conservative 0; Mismatches 209; Indels 0; Gaps 0;
QY 320 TCCTCCAGGAGATTGCGCGTCCACCATCCCTTACGAGAACCGGGAGCTCTCGAGGAAGG 379
Db 155 TCGCCAGCTGCTGCGCGTCCAGGAGCGCGGCCATTGCGCTCCAGGCGACGCG 214
QY 380 TGGGAGCGGTCTTCTCATGTCTCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCAACC 439
Db 215 CGCGCGGAGGAAAGCGAGGCGCGCAAGAGCAGCTCAACTGTTCAGCGCGCGCGG 274
QY 440 GCGAGACCGTCCACCGCTGGAAGAGTCTCTTGAGAAAAGGGCTGTGGCCACCGACG 499
Db 275 CGCGCGGCTGACCGGTTTCGAGGAGGTTCTGTCGCAACGTCTCGAGCGCGCCACG 334
QY 500 TCCTTCACCAACCGTCAACGGGAGCGCGGCCCATCGGCACCTTTGGCGCTCGCGC 559
Db 335 GGTGCGCGGAGCGCGGACCGCGCGTGCAGATCGCGGCAACTTGGCGCGCTCGGG 394
QY 560 TGAGGCCAGGAAAGCAGGCTCACTCCCTGGAGACTACTACTACCCCTGGAGGAACCTCG 619
Db 395 AGAGCGCGCGGCTGACGAGCTCCCGCTTCGCGCGGATCCCGCCCTTCATCGAGGGG 454
QY 620 CCTAGACATGCCAACGCGGTGCTCTCTTCAACTGGGTCAAGGCTCAAGGACCAACG 679
Db 455 TCTAGCGCGGAGCGCGGCAACCGCTGCTTCGACCGCTCGCGGGGACCACTCTTCG 514
QY 690 GAATCGCGCGGACCTTGGACGTGTG 705
Db 515 ACGGCGAGCGCATGTTGACACGCGCTG 540
Search completed: March 17, 2004, 21:47:59
Job time : 255.717 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 21:15:22 ; Search time 47.9049 Seconds
(without alignments)
11885.630 Million cell updates/sec

Title: US-09-664-186-4

Perfect score: 1026

Sequence: 1 gtgaagacgaaacacatt.....ccctcaaggagcctcctga 1026

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1026	100.0	1026	3	US-09-134-246-4
2	1026	100.0	5849	3	US-09-134-246-6
3	51.8	5.0	1551	4	US-09-252-991A-5309
4	51.8	5.0	1563	4	US-09-252-991A-15934
5	51.8	5.0	1953	4	US-09-252-991A-15763
6	51.8	5.0	3351	4	US-09-252-991A-15871
7	51.8	5.0	4158	4	US-09-252-991A-5348
8	51.8	5.0	4953	4	US-09-252-991A-5227
9	49.6	4.8	4403765	3	US-09-103-840A-2
10	49.6	4.8	4411529	3	US-09-103-840A-1
11	46.8	4.6	2460	4	US-09-252-991A-2998
12	46.8	4.6	2856	4	US-09-252-991A-2869
13	46.8	4.6	3387	4	US-09-252-991A-3101
14	46.2	4.5	1473	4	US-09-152-060-43
15	46	4.5	1572	4	US-09-252-991A-8725
16	46	4.5	1929	4	US-09-252-991A-8827
17	46	4.5	2079	4	US-09-252-991A-9222
18	45.8	4.5	960	4	US-09-252-991A-6990
19	45.8	4.5	1275	4	US-09-252-991A-7059
20	45.8	4.5	1926	4	US-09-249-585A-4
21	45.8	4.5	1931	2	US-08-130-114-2
22	45.8	4.5	2187	4	US-09-252-991A-7042
23	45.6	4.4	984	4	US-09-252-991A-10162
24	45.6	4.4	1491	4	US-09-252-991A-10097
25	45.6	4.4	2668	2	US-08-461-775-11
26	45.6	4.4	2668	3	US-09-031-606-11
27	45.4	4.4	1758	4	US-09-252-991A-490

RESULT 1

US-09-134-246-4

; Sequence 4, Application US/09134246B

; Patent No. 6207377

; GENERAL INFORMATION:

; APPLICANT: Wayne, Jay

; APPLICANT: Xu, Shuang-Yong

; TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
; TITLE OF INVENTION: Vectors And Identification Of Two Thermus Plasmid

; TITLE OF INVENTION: Replication Origins

; FILE REFERENCE: Thermus Shuttle Vector

; CURRENT APPLICATION NUMBER: US/09/134,246B

; CURRENT FILING DATE: 1998-08-14

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 1026

; TYPE: DNA

; ORGANISM: Thermus sp.

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1023)

US-09-134-246-4

Query Match 100.0%; Score 1026; DB 3; Length 1026;

Best Local Similarity 100.0%; Pred. No. 8.3e-241;

Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGAACGAAACGAAACCTCTTTGAAGAGCTTTAGAGGCTTTAGAGAAACCCACGAC 60

Db 1 GTGAACGAAACGAAACCTCTTTGAAGAGCTTTAGAGGCTTTAGAGAAACCCACGAC 60

Qy 61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTTGGCCACCGAC 120

Db 61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTTGGCCACCGAC 120

Qy 121 CCCCCCTCAGATGGAGGTCCGAAATCCCTCGGAGGGCTTTACATACCAAAAGAG 180

Db 121 CCCCCCTCAGATGGAGGTCCGAAATCCCTCGGAGGGCTTTACATACCAAAAGAG 180

Qy 181 GCACCTTAGGATTCCTTACCCGAGAAAGACCATGAGGCTTTCTTTCTCTCTGTGGGGCC 240

Db 181 GCACCTTAGGATTCCTTACCCGAGAAAGACCATGAGGCTTTCTTTCTCTCTGTGGGGCC 240

Qy 241 CCCCCCTATACCCAGCTGAACCCCGTTGGGAATGTATGTCAAGCCCTCAGAGCGGG 300

Db 241 CCCCCCTATACCCAGCTGAACCCCGTTGGGAATGTATGTCAAGCCCTCAGAGCGGG 300

Qy 301 CCTCAGAGCTTCTGGAACCTCTCCAGGAGATTGCCCGCTCCACCATCCCTACGCAAC 360

Db 301 CCTCAGAGCTTCTGGAACCTCTCCAGGAGATTGCCCGCTCCACCATCCCTACGCAAC 360

Sequence 568, App
Sequence 11487, A
Sequence 11405, A
Sequence 6208, Ap
Sequence 6341, Ap
Sequence 11988, A
Sequence 11987, A
Sequence 12025, A
Sequence 76, Appl
Sequence 4221, Ap
Sequence 4664, Ap
Sequence 4726, Ap
Sequence 175, App
Sequence 4686, Ap
Sequence 4757, Ap
Sequence 1, Appli
Sequence 11385, A
Sequence 42, Appl

ALIGNMENTS

RESULT 1

US-09-134-246-4

; Sequence 4, Application US/09134246B

; Patent No. 6207377

; GENERAL INFORMATION:

; APPLICANT: Wayne, Jay

; APPLICANT: Xu, Shuang-Yong

; TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle
; TITLE OF INVENTION: Vectors And Identification Of Two Thermus Plasmid

; TITLE OF INVENTION: Replication Origins

; FILE REFERENCE: Thermus Shuttle Vector

; CURRENT APPLICATION NUMBER: US/09/134,246B

; CURRENT FILING DATE: 1998-08-14

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 1026

; TYPE: DNA

; ORGANISM: Thermus sp.

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1023)

US-09-134-246-4

Query Match 100.0%; Score 1026; DB 3; Length 1026;

Best Local Similarity 100.0%; Pred. No. 8.3e-241;

Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGAACGAAACGAAACCTCTTTGAAGAGCTTTAGAGGCTTTAGAGAAACCCACGAC 60

Db 1 GTGAACGAAACGAAACCTCTTTGAAGAGCTTTAGAGGCTTTAGAGAAACCCACGAC 60

Qy 61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTTGGCCACCGAC 120

Db 61 AACACCGATGCCACTAGGGGGTCAGATAGGGGGTCAGAGGACTTCTTTGGCCACCGAC 120

Qy 121 CCCCCCTCAGATGGAGGTCCGAAATCCCTCGGAGGGCTTTACATACCAAAAGAG 180

Db 121 CCCCCCTCAGATGGAGGTCCGAAATCCCTCGGAGGGCTTTACATACCAAAAGAG 180

Qy 181 GCACCTTAGGATTCCTTACCCGAGAAAGACCATGAGGCTTTCTTTCTCTCTGTGGGGCC 240

Db 181 GCACCTTAGGATTCCTTACCCGAGAAAGACCATGAGGCTTTCTTTCTCTCTGTGGGGCC 240

Qy 241 CCCCCCTATACCCAGCTGAACCCCGTTGGGAATGTATGTCAAGCCCTCAGAGCGGG 300

Db 241 CCCCCCTATACCCAGCTGAACCCCGTTGGGAATGTATGTCAAGCCCTCAGAGCGGG 300

Qy 301 CCTCAGAGCTTCTGGAACCTCTCCAGGAGATTGCCCGCTCCACCATCCCTACGCAAC 360

Db 301 CCTCAGAGCTTCTGGAACCTCTCCAGGAGATTGCCCGCTCCACCATCCCTACGCAAC 360

Sequence 568, App
Sequence 11487, A
Sequence 11405, A
Sequence 6208, Ap
Sequence 6341, Ap
Sequence 11988, A
Sequence 11987, A
Sequence 12025, A
Sequence 76, Appl
Sequence 4221, Ap
Sequence 4664, Ap
Sequence 4726, Ap
Sequence 175, App
Sequence 4686, Ap
Sequence 4757, Ap
Sequence 1, Appli
Sequence 11385, A
Sequence 42, Appl

Db	301	CCTCAGAGCTTCTGGAACTCTCCAGGAGATTGCCCGCTCCACCATCCCTACGGCAAC	360	Best Local Similarity 100.0%; Pred. No. 1.3e-240; Matches 1026; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	361	CGGAGCTCTGAGGAAGTGGGAGCGTCTCTTCATGCTTCATGCTCCCTCGAGATGTTGCC	420	
Db	361	CGGAGCTCTGAGGAAGTGGGAGCGTCTCTTCATGCTTCATGCTCCCTCGAGATGTTGCC	420	
Qy	421	CTCAACTCTGGGGTCACTCCGGCAGACCGTCCACGCTCGGAGAAAGTCTCTTGAGAAAAA	480	
Db	421	CTCAACTCTGGGGTCACTCCGGCAGACCGTCCACGCTCGGAGAAAGTCTCTTGAGAAAAA	480	
Qy	481	GGCTTGGTGGCCACCGAAGTCTCTTCAACAAACCGTCAACGGGAGCGCGGCGCCATCGC	540	
Db	481	GGCTTGGTGGCCACCGAAGTCTCTTCAACAAACCGTCAACGGGAGCGCGGCGCCATCGC	540	
Qy	541	ACCTTTGGGGCTCGGCTGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	600	
Db	541	ACCTTTGGGGCTCGGCTGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	600	
Qy	601	TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAAAGCGGCTCTCTCTCAACTGGGTC	660	
Db	601	TACCCCTGGAGGAACCTCGCCCTAGACATGGCCAAAGCGGCTCTCTCTCAACTGGGTC	660	
Qy	661	AGGCTTACAGGACACGAGATCCGCCACCCCTGAGGCTGCTGCTCTGAGGCTCAG	720	
Db	661	AGGCTTACAGGACACGAGATCCGCCACCCCTGAGGCTGCTGCTCTGAGGCTCAG	720	
Qy	721	GGGAAAGGCTGATGCCCAACACCAAGACCGTGGCGCTTGAACCTGGGCTCATCTGGTC	780	
Db	721	GGGAAAGGCTGATGCCCAACACCAAGACCGTGGCGCTTGAACCTGGGCTCATCTGGTC	780	
Qy	781	CTCCCGAGGTGAGGCTTCCAAACTCCCGGCGCTTATCACCCTCATTTGCTACGAT	840	
Db	781	CTCCCGAGGTGAGGCTTCCAAACTCCCGGCGCTTATCACCCTCATTTGCTACGAT	840	
Qy	841	GCGATCTCTAGATGACCGTGTCTCAAGACGTTCTATGACGGCTGTGCTGGGCTGTG	900	
Db	841	GCGATCTCTAGATGACCGTGTCTCAAGACGTTCTATGACGGCTGTGCTGGGCTGTG	900	
Qy	901	GCAGGGTGAATCCCGCGCAATATCTATTGGCGTCTTAATCGGGTTATCGAGAT	960	
Db	901	GCAGGGTGAATCCCGCGCAATATCTATTGGCGTCTTAATCGGGTTATCGAGAT	960	
Qy	961	TACAGGATGGCCATCTGACACGACCGGAGGAGTACCTAGTAGAGACCTCAAGAGGCC	1020	
Db	961	TACAGGATGGCCATCTGACACGACCGGAGGAGTACCTAGTAGAGACCTCAAGAGGCC	1020	
Qy	1021	TCCTGA 1026		
Db	1021	TCCTGA 1026		
RESULT 2				
US-09-134-246-6/c				
; Sequence 6, Application US/09134246B				
; Patent No. 6207377				
; GENERAL INFORMATION:				
; APPLICANT: Wayne, Jay				
; TITLE OF INVENTION: Method For Construction Of Thermus-E. coli Shuttle				
; TITLE OF INVENTION: Vectors And Identification Of Two Thermus Plasmid				
; FILE REFERENCE: Thermus Shuttle Vector				
; CURRENT APPLICATION NUMBER: US/09/134, 246B				
; NUMBER OF SEQ ID NOS: 30				
; SOFTWARE: PatentIn Ver. 2.0				
; SEQ ID NO 6				
; LENGTH: 5849				
; TYPE: DNA				
; ORGANISM: Thermus sp.				
US-09-134-246-6				
Query Match 100.0%; Score 1026; DB 3; Length 5849;				

308 AGTTCTGGAATCTCTCCAGGAGATTGCCCGCTCCACCATCCCTTACGGCAACCGGAGC 367
1412 AGCGCTGGAGGCTTCCAGTGGAGCGGGCGACATCGTCTGAAGGCGCGCAACTGA 1353
368 TCTGGAGGAGTGGGACGGTCTCTTCATGGTCCCTTGAGATGTTGGCCCTCAACC 427
1352 AGTCGGCAACCTCGACAGTTTCGACCTGATCACCGGAGCGCAAGCTCAACGCCAAGC 1293
428 TGGGGGTACCCGGGAGACCGTCCACGCTTGGAAAGAGTCTTGGAAAGAGGCGCTGG 487
1292 TCTACGCGAAGAACCTCAACATCGTCCACCGCGCGCAACGACGCTCCAGGCGCGACGCTGC 1233
488 TGGCCACCGACGTCCTTCCACCAACCGTCAACGGGAGCGCGGCCCTCGSCACCCCTT 547
1232 AGCCACAGCGCGCGCGCGCGGATGGAGCGAAGCCACAGTGGCGATCGACAGCTCGG 1173
548 GGGCCCTCCGGTGTAGGCCAGGGAAGCCAGGCTCACCTGGACGACTACATCTTACCCT 607
1172 CGCTGGCGGATGTACGCGCGGGCGATCCGCTGTGTGGCACCGCAACAGGCGGTGGGG 1113
608 GGAGGAACCTCGCCTAGACATGGCCAAACGGCGTCTCTTCAACTGGTCAAGGCT 667
1112 TGGGCTGCGCGGACATGGCGCGCGCGGCGGACATCCGCATCGACGCCAGCGCA 1053
668 ACCAGGACACGGAATCCGCGCCACCCCTGGAGCTGTGGTCTCTTGGGCTCAGGGGAAA 727
1052 AGTGAAGCTGGCCAGGCTCCAGCCAGGCGGACCTGAAGATCGCGCCAGGCGGTGG 993
728 GGGTGTATGCCCAACACAGACAGCTGGCGGCTGTACCTGGGCTCATCTGG 778
992 AGCTGAACGCGAGACCTACGCGGTGGCAGCGCGGAGATCCGCGAGCGG 942

RESULT 6
US-09-252-991A-15871
; Sequence 15871, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15871
; LENGTH: 3351
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15871

Query Match 5.0%; Score 51.8; DB 4; Length 3351;
Best Local Similarity 44.4%; Pred. No. 0.0021;
Matches 209; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

308 AGTTCTGGAATCTCTCCAGGAGATTGCCCGCTCCACCATCCCTTACGGCAACCGGAGC 367
947 AGCGCTGGAGGCTTCCAGTGGAGCGGGCGACATCGTCTGAAGGCGCGCAACTGA 1006
368 TCTGGAGGAGTGGGACGGTCTTTCATGGTCCCTTGAGATGTTGGCCCTCAACC 427
1007 ACCTGCGCAACCTCGAAGCTTTCGACCTGATCAACCGCGAGCGCAAGCTCAACGCCAAGC 1066
428 TGGGGGTACCCGGCAGACGCTCCAGCTTGAAGAGTCTTGAAGAAAAGGGCGCTGG 487
1067 TCTACGGAAGAACCTCAACATCGTACCGGCGGCAACGACGCTCCAGGCGGACAGCTGC 1126
488 TGCCCAACCGAGTCTCTTCAACCAACCGGTCAACGGGAGCGCGGCCATTCGGCACCCCTTT 547

1127 AGCCACGCGCGCGCGCGATGGCAGGAGCAACAGCTGGCGATCGACAGCTCGG 1186
548 GGGCGCTCCGGCTGAGCGCAGGGAAGCCAGCTCACCTGGAGCACTACTTACCCCT 607
1187 CGTGGGCGGGATGTACGCGGGGCGATCCGCTGTGTGGCACCAGGAGGCGTGGGG 1246
608 GGAGGAACCTCGCCCTAGACATGGCCAAAGGGCGTGTCTCTCTTCAACTGGGTCAAGGCT 667
1247 TGGCGCTGGCGCGGACATGGCGCGCGCAGCGGCGGACATCCGATCGACGCCAGCGCA 1306
668 ACCAGGACACGGAATCCGCGCCACCTCGACGTGTGTCTCTTGGGCTCAGGGGAAA 727
1307 AGTGAAGCTTGGCCCGAGGCTCCAGCCAGGCGACCTGAAGATCGCGGCCAGGCGGTGG 1366
728 GGGTGTATGCCCAACACCAAGACCGTGGCGCTTGACCTGGGCTCATCTGG 778
1367 AGCTGAACGCGCAAGCACTACGCGCGGTGGCAGCGCGGATCCGACGGCGG 1417

RESULT 7
US-09-252-991A-5348
; Sequence 5348, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5348
; LENGTH: 4158
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5348

Query Match 5.0%; Score 51.8; DB 4; Length 4158;
Best Local Similarity 44.4%; Pred. No. 0.0022;
Matches 209; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

308 AGTTCTGGAATCTCTCCAGGAGATTGCCCGCTCCACCATCCCTTACGGCAACCGGAGC 367
587 AGCGCTTGGAGCGCTTCCAGGTGAGCGCGGCGACATCGTCTGAAGGCGCGCAACTGA 646
368 TCTGGAGGAGTGGGAGCGGTCTTTCATGGTCCCTTGGAGATGTTGGCCCTCAACC 427
647 AGCTCGCAACCTCGAAGCACTTCGACCTGATCACCGCAGCGCAAGCTCAACGCCAAGC 706
428 TGGGGGTACCCGGCAGACCGCTCCAGCTTGAAGAGTCTTGTGAGAAAAGGGCGCTGG 487
707 TCTACGCGAAGAACCTCAACATCGTACCGCGCGCAACGACGCTCCAGGCGCGACGCTGC 766
488 TGGCCACCGAGCTCTTCAACCAACCGTCAACGGGAGCGCGGCCATCGGCACCCCTTT 547
767 AGGCCACCGCGCGCGCGCGATCGCGCTGTGGCAGGAGCCACAGCTGGCGATCGACGCTCGG 826
548 GGGCGCTCCGGCTGAGCGCCAGGGAAGCCAGGCTCACCTTGAAGCAAGTCTTACCTTCAAGGCTCAAGGCT 607
827 CGCTGGGCGGATGTACCGCGGGCGGATCCGCTGTGTGGCACCAGCAGCGGCTGGGG 886
608 GGAGGAACCTCGCCCTAGACATGGCCAAACGGCGTGTCTCTTCAACTGGGTCAAGGCT 667
887 TGGCGCTGGCGGACATGGCGCGCGGCGGCGGCGGACATCCGCGATCGACGCCAGCGCA 946
668 ACCAGGACACGGAATCCGCGCCACCCCTGGACGCTGTGTCTCTTGGGCTCAGGGGAAA 727
947 AGCTGAGCTTGGCCCGAGGCTTCCAGCGCGGCGGACCTGAAGATCGCGGCCAGGCGGTGG 1006

QY 728 GGGTGTGCCCCAACCAACCAAGACCGTGGCCCTTACCTGGGCCCTCATCTCG 778
Db 1007 AGCTGAACGGCAAGACCTACCGCGTGGCAGCGCGAGATCCGACGCGCG 1057

RESULT 8
US-09-252-991A-5227/c
; Sequence 5227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 5227
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-5227

Query Match 5.0%; Score 51.8; DB 4; Length 4953;
Best Local Similarity 44.4%; Pred. No. 0.0023;
Matches 209; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

QY 308 AGCTTCTGGAACCTCTCCAGAGATTGCCCGCTCCACATCCCTACGCGCAACCGGAGC 367
Db 4412 AGCCCTTGGAGCGTTTCCAGTGGACGCGCGGACATCTGTCGAGGCGCGCACTGA 4353

QY 368 TCTGAGGAAGGTGGGACCGTGTCTTCTATGTTCCCTCGGAGATTTGGCCCTCAACC 427
Db 4352 ACGTCCGCAACCTCGAACAGTTCCAGCTGATCACCCTGCGGCAAGCTCAACGCCAAGC 4293

QY 428 TGGGGTTCACCCGACACCGTCCACCTCGAAGAGTCTTGAGAAAAGGGCTGG 487
Db 4292 TCTACGCGAAGACCTCAACATGTCTACCGCGCGCAACGACGTCCAGGCGCAGGCTGC 4233

QY 488 TGGCCACCGAGCTCTTTCACAAACCGTCAACGGGAGCGCCGGCCATCGGCACCCCTTT 547
Db 4232 AGGCCACGCGCGCGCGCGATGGCAGCGAGAGCCACAGCTGGCGATCGACGCTCGG 4173

QY 548 GGGCCGTCCGCTGAGGCCAGGGAAGCAGGCTCACCTCGGAGACTATCTACCCCT 607
Db 4172 CGTGGCGGGGATGTACGCGGGGCGATCCGCTGTGTCGACCGAGCGGCGTGGGG 4113

QY 608 GGAGAACCTCGCCCTAGACATGCCAAACGCGTGTCTCTTCAACTGGGTCAAGGCCT 667
Db 4112 TGGGCTGGCGCGGACATGGCCGCGAGCGCGCGACATCCGATCGACGCCAGCGGCA 4053

QY 668 ACCAGACACCGGAATCCGCCCAACCGTGGTGTGCTCTCTGGGTTCAGGGGAAA 727
Db 4052 AGCTGAGCTGGCCAGCGCTCCAGCAGCGGACCTGGAAGATCGCGGCCAGCGCTGG 3993

QY 728 GGGTGTGCCCCAACCAACCAAGACCGTGGCCCTTACCTGGGCCCTCATCTCG 778
Db 3992 AGCTGAACGGCAAGACCTACCGCGTGGCAGCGCGGAGATCCCGACGCGG 3942

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 4.8%; Score 49.6; DB 3; Length 4403765;
Best Local Similarity 45.7%; Pred. No. 0.053;
Matches 212; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

QY 391 GTCTTCATGTCCCTCGAGATGTGGCCCTCAACCTGGGGGTCAACCGGCGAGACCTC 450
Db 2780667 GCCTTCTGGCCACCGCGGCCACCGTGGCATCTCTGCAACAGGAACCGCGCTGAACGAG 2780608

QY 451 CACGCTGGAAGAAGTCTCTTGAAGAAAGGGCCCTGTGGCCACCGAGCTCTTCAACCA 510
Db 2780607 GACAAGACGTTTCGGCGCAATGTGGAAGAGGGCATGCGGGGACATCAAGATCAAGCTCGAC 2780548

QY 511 ACCGTCAACGGGAGCGCGGCCCATCGGCACCCCTTTGGGGCGCTCGGCTGAGGCCAGG 570
Db 2780547 CGCTTCAACGAGGTCCGCGAATTGATGCCACCGACTACCGACGAGCTGATGGAAGAG 2780488

QY 571 AAAGCAGGCTCACCTCGAGACTACATCTACCTCGAGGAACCTCGCCCTAGACATG 630
Db 2780487 ATGGTTCGGCTGCAAGAGGAATCGACCAACCGCGCTGGGACCTCGACGCGCAGCTC 2780428

QY 631 GCCAACGGCGTCTCTCTTCAACTTGGGTCAAGGCTTACGAGGACCGAGATCCGCCCC 690
Db 2780427 GAGCAGGCAATGGATGCGCTGTCGCGGCCGAC--GAGCCGTTAACCAACCTA 2780371

QY 691 ACCGTGAGCTGTGGTCTCTTGGGCTCAGGGGAAAGGGTGTATGCCCAACCAAGACC 750
Db 2780370 TCCGTTGGCGAGCTGCGCGGGTGGCGCTGTGCAAACTGCTGTTCGCAAAACCCGACCTG 2780311

QY 751 GTGGCGGTGACCTGGGCTCATCTCTGCTCTCCCGAGGTGGAGCTTCCAACTCCCG 810
Db 2780310 TTGTTGTCGAGAGCCGACCAACCACTAGACGCGGAAAGTGTGAGTGGCTCGAAGAG 2780251

QY 811 GCCCTTATCACCCCTCATCTGCTACGTACATTCGCGATCTCTCTAGA 854
Db 2780250 CATCTGGCCAGCTACCCCGTGGATCTCTGGCGGTCAACCCACGA 2780207

RESULT 10
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529

TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 4.8%; Score 49.6; DB 3; Length 4411529;
Best Local Similarity 45.7%; Pred. No. 0.053; Indels 3; Gaps 1;
Matches 212; Conservative 0; Mismatches 249;

QY 391 GTCTTCATGTCCTCCCTCGAGATGTTGGCCCTCAACCTGGGGTCAACCGGACAGCGTC 450
DB 2783857 GCCTTCCTGGCCACCGCGCACCGTGGCATCTCTGCAACAGGAACCGCCGCTGAACGAG 2783798

QY 451 CAGGCTGGAAGAAGTCTTGAGAAAAGGCGCTGGTGGCCACCGAGCTCTTCACCAA 510
DB 2783797 GACAAAGACGTTTCGGGCAATGTGAAGAGGGCATGGGGGACATCAAGATCAAGCTGCAC 2783738

QY 511 ACCGTCAACGGGGAGCGCGGCCCATCGGCACCTTTGGGGCGTCCGGCTGAGGCCAGGG 570
DB 2783737 CGCTTCAAGAGGTCCCGCAATGATGGCCACCGACTACACCGAGAGCTGATGGAAG 2783678

QY 571 AAAGCAGGCTCACCTTGAGACGACTACATCTACCTCGAGGAACCTCGCCCTAGACATG 630
DB 2783677 ATGGTTCGCTCAAGAGGAAGTGAACCAACCGCGCTGGACCTCGACGCGAGCTC 2783618

QY 631 GCCAAGCGGTCTCTCTCTTCAACTGGGTCAAGGCTTACCAGGACCAAGGAATCGGCCCC 690
DB 2783617 GAGCAGGCATGATGGCTGCGCTGTCCGCGGCCGAC---GAGCGGTAAACCACTA 2783561

QY 691 ACCCTGAGCTCTGCTCTCTGCTCTGAGGCTCAGGGAAGGAGTATGCCCAACCAAGACC 750
DB 2783560 TCCGCTGGGAGCGCTCGCGGGTGGCTGTGCAAACTGCTGTGTCCAAACCGGACCTG 2783501

QY 751 GTGGCGGTGACTGGGCTCATCTGCTCTCCCGAGGTGGAGCGTTCCAAACTCCCG 810
DB 2783500 TTGTTGCTCGAGAGCGGACCAACCACTTAGACGCGGAAGTGTGCACTGGCTCGAACAG 2783441

QY 811 GGCCTTATCACCTCATCTGCTACGTACATTTGCGGATCTCTTAGA 854
DB 2783440 CATCTGGCAGCTACCCGGTGGATCTCTGGGGTCAACCCAG 2783397

RESULT 11
US-09-252-991A-2998
; Sequence 2998, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2998
; LENGTH: 2460
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2998

Query Match 4.6%; Score 46.8; DB 4; Length 2460;
Best Local Similarity 46.4%; Pred. No. 0.032;
Matches 153; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 356 GCNACCGGAGCTCTGAGGAAGGTGGGACGGTCTTTCATGTCCTCCCTCGGAGATGT 415
DB 1500 GCAACCGCTGCTGAGGATTTCCGCACTGTACGGCATTCGAGCATCCGACGCTCCG 1559

QY 416 TGGCCCTCAACCTGGGGTCAACCGGACAGCGTCCACCGCTGGGAAGGTCCTTGAGA 475

Db 1560 GGGAGACCCCTGGAGCAGTGGAGGCGAGCTCATGCCGAGGATCACGACCGCTGC 1619

QY 476 AAAAGGCTGTGTGGCCACCGAGCTCTTACCACCAACGTCACCGGAGCGCCGGGCA 535

Db 1620 TGGCGAGCTGGAGCGCGCGCTCAGCGGTTTCAGAAAATTCGAGTTCCGATATCG 1679

QY 536 TGGGACCCCTTTGGGCGCTCGGCTGAGGCCAGGAAAGCCAGGCTCACCTTGGAGCACT 595

Db 1680 TCGCCCCCAACCGGAGTTTCGCCACCTGCGGCGCATGCCACGCTGACGCGGACGAGG 1739

QY 596 ACATCTACCCCTGGAGAACCTTCGCCCTTAGACATGGCCAAACGCGTGTCTCTCTCACT 855

Db 1740 ACAACCGTCCGCTGCGATGATCGGCATCAACAGCGACATCACCGAGATCCGACCCCTGG 1799

QY 656 GGGTCAAGCCCTACGAGGACCGGAATCC 685

Db 1800 CGGAAACCCCTGCACGAGGAAAGGAACGCC 1829

RESULT 12
US-09-252-991A-2869/c
; Sequence 2869, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2869
; LENGTH: 2856
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2869

Query Match 4.6%; Score 46.8; DB 4; Length 2856;
Best Local Similarity 46.4%; Pred. No. 0.033;
Matches 153; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 356 GCAACCGGAGCTCTGAGGAAGGTGGGACGGTCTTTCATGTCCTCCCTCGGAGATGT 415
DB 2027 GCAACCGCTGCACTGGGATTTCCGCACTGTTCGACCTGTACGCGCATCCGACGGTCCCG 1968

QY 416 TGGCCCTCAACCTGGGGTCAACCGGACAGCGTCCACCGCTGGGAAGGTCCTTGAGA 475

Db 1967 GGGAGACCGCGCTGGAGCAGTGAAGCGGAGCTGTATGCCGAGGATCAGACCGCTGC 1908

QY 476 AAAAGGCTGTGTGGCCACCGAGCTCTTACCAAAACCTTCAACCGGAGCGCCGGGCA 535

Db 1907 TGGCGAGCTGGAGCGCGCGCTCAGCGGTTCAGAAATTCGACTCGGAGTTCCGATATCG 1848

QY 536 TCGGCAACCTTTGGGCGCTCGGCTGAGCGGAGGAAAGCCAGGCTCACCTTGGAGCACT 595

Db 1847 TCGCCCCCAACCGGAGTTTCGCCACCTCGGGCGCATCCGACGCTGACGCGGACGAG 1788

QY 596 ACATCTACCCCTGGAGAACCTTCGCCCTTAGACATGGCCAAACGCGGCTCTCTCTCACT 655

Db 1787 ACAACCGTCCGCTGCGATGATCGGCATCAACAGCGACATCACCGAGATCCGACCCCTGG 1728

QY 656 GGGTCAAGCCTTACGAGGACCGGAATCC 685

Db 1727 CGGAAACCCCTGCACGAGGAAAGGAACGCC 1698

RESULT 13
US-09-252-991A-3101

; Sequence 3101, Application US/09252391A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 3101
 ; LENGTH: 3387
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-3101

Query Match 4.6%; Score 46.8; DB 4; Length 3387;
 Best Local Similarity 46.4%; Pred. No. 0.035;
 Matches 153; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 356 GCAACCGGGAGCTCTCGAGGAGGTGGGACGGTCTTTCATGTCCTCCCTCGAGATGT 415
 DB 1364 GCAACCGCTGCACCTGGATTTCGGCATGTTGCACCTGTACGGCATCCGACGGGTCCCG 1423
 QY 416 TGGCCCTCAACCTGGGGGTCAACGGGAGACCTGCACGGCTGGAAGAGGTCTTTGAGA 475
 DB 1424 GGAGACCGCGCTGGAGCACTGGAAGCGAGCTGTCATGTCGGAGGATCAAGACCGGTGC 1483
 QY 476 AAAAGGGCTTGGTGCCACCGACGCTCTTCCACAAACCGTCAACGGGGAGCGCGGGCCA 535
 DB 1484 TGGCGCAGCTGGAGCGCGCGCTCAGCGGTTGCAGAAATTCGACTCGAGTTCGGTATCG 1543
 QY 536 TCGGCACCTTTTGGGCGGTTCGGCTGAGGCCAGGAAAGCGACGGCTCACCTCGAGACT 595
 DB 1544 TCGCGCCCAACCGCGAGTTTCGCACCTCGGGGCGATCGCCAGCTCAGCGCGAGG 1603
 QY 596 ACATCTACCTCGAGGAACCTCGCCTAGACATGCCACGGCGGTCTCTCTTCAACT 655
 DB 1604 ACAACGTCGCGTGGGATGATCGGATCAACAGGACATACCGAGATCCGACCTGG 1663
 QY 656 GGGTCAGGCTTACCAGGACCGAGATCC 685
 DB 1664 CGGAACCTCGACGAGGAAGGAAGCC 1693

RESULT 14
 US-09-152-060-43/c
 ; Sequence 43, Application US/09152060
 ; Patent No. 6448230
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: 28 Human Secreted Proteins
 ; FILE REFERENCE: P2003P1.US
 ; CURRENT APPLICATION NUMBER: US/09/152,060
 ; CURRENT FILING DATE: 1998-09-11
 ; EARLIER APPLICATION NUMBER: PCT/US98/04858
 ; EARLIER FILING DATE: 1998-03-12
 ; EARLIER APPLICATION NUMBER: 60/040,762
 ; EARLIER FILING DATE: 1997-03-14
 ; EARLIER APPLICATION NUMBER: 60/040,710
 ; EARLIER FILING DATE: 1997-03-14
 ; EARLIER APPLICATION NUMBER: 60/050,934
 ; EARLIER FILING DATE: 1997-05-30
 ; EARLIER APPLICATION NUMBER: 60/048,100
 ; EARLIER FILING DATE: 1997-05-30
 ; EARLIER APPLICATION NUMBER: 60/048,357
 ; EARLIER FILING DATE: 1997-05-30
 ; EARLIER APPLICATION NUMBER: 60/048,189
 ; EARLIER FILING DATE: 1997-05-30

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 18, 2004, 02:39:53 ; Search time 235.943 Seconds
(without alignments)

16085.819 Million cell updates/sec

Title: US-09-664-186-4

Perfect score: 1026

Sequence: 1 Gtaagaacgaaacaccc.....ccctcaaggagcctcctga 1026

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

```
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
c 1	53.8	5.2	9025608	14	US-10-156-761-1
2	53.2	5.2	1755	15	Sequence 1, Appli
3	51.6	5.0	1815	9	Sequence 721, App
4	50.8	5.0	1018	12	Sequence 13, Appli
5	50.8	5.0	1077	14	Sequence 35941, A
6	50.8	5.0	9025608	14	Sequence 6573, Ap
7	50	4.9	675	14	Sequence 1, Appli
8	49.8	4.9	1389	14	Sequence 6076, Ap
9	49.6	4.8	1674	12	Sequence 5431, Ap
10	49.6	4.8	1677	12	Sequence 26289, A
11	49.4	4.8	1803	12	Sequence 28540, A
12	49.2	4.8	1045	15	Sequence 32815, A
13	48.8	4.8	1803	12	Sequence 42225, A
14	48.6	4.7	960	15	Sequence 26418, A
15	48.6	4.7	9369	14	Sequence 34210, A
					Sequence 190, App

48.6	4.7	9369	14	US-10-237-551-190	Sequence 130, App
48.6	4.7	9369	14	US-10-237-551-247	Sequence 247, App
48.6	4.7	154746	10	US-09-827-688-8	Sequence 8, Appli
48	4.7	1977	14	US-10-355-430-17	Sequence 17, Appl
47.8	4.7	963	12	US-10-282-122A-11300	Sequence 11900, A
47.6	4.6	669	12	US-10-425-114-34873	Sequence 34873, A
47.6	4.6	864	12	US-10-425-114-35445	Sequence 35445, A
47.6	4.6	1527	14	US-10-156-761-6366	Sequence 6366, Ap
47.6	4.6	1926	14	US-10-156-761-2557	Sequence 2557, Ap
47.2	4.6	969	14	US-10-156-761-111	Sequence 111, App
46.8	4.6	1224	14	US-10-156-761-4891	Sequence 4891, Ap
46.8	4.6	1605	15	US-10-260-338-908	Sequence 908, App
46.8	4.6	2223	14	US-10-156-761-3605	Sequence 3605, Ap
46.6	4.5	1071	15	US-10-369-493-42984	Sequence 42984, A
46.6	4.5	10665	14	US-10-160-758-1	Sequence 1, Appli
46.6	4.5	10065	15	US-10-085-117-114	Sequence 114, App
46.6	4.5	10455	14	US-10-160-758-4	Sequence 4, Appli
46.6	4.5	11073	14	US-10-160-758-2	Sequence 2, Appli
46.6	4.5	11073	15	US-10-085-117-113	Sequence 13514, A
46.4	4.5	1381	12	US-10-425-114-13514	Sequence 13514, A
46.2	4.5	1414	12	US-10-425-114-14100	Sequence 14100, A
46.2	4.5	1473	9	US-09-853-161-43	Sequence 43, Appl
46.2	4.5	1473	9	US-09-852-659A-43	Sequence 43, Appl
46.2	4.5	1473	9	US-09-852-797-43	Sequence 2448, Ap
46.2	4.5	2892	14	US-10-156-761-2448	Sequence 52912, A
46	4.5	808	12	US-10-424-599-52912	Sequence 49502, A
46	4.5	1263	15	US-10-369-493-40502	Sequence 1960, Ap
46	4.5	1418	12	US-10-425-114-1960	Sequence 34918, A
46	4.5	1532	12	US-10-425-114-34918	Sequence 28403, A
46	4.5	1964	12	US-10-425-114-28403	

ALIGNMENTS

RESULT 1

```
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
```

Query Match 5.2%; Score 53.8; DB 14; Length 9025608;
Best Local Similarity 45.0%; Pred No. 0.00014;
Matches 202; Conservative 0; Mismatches 247; Indels 0; Gaps 0;

QY 261 ACCCCCGGTGGGAATGTATGTCAAGCCGCTCAGACCGGCTCAGAGCTTCTGGAAT 320

DB 4922513 ACCCGTCCCGCGCAGGGGTATCGTGGCGCTCATGACCCGCTCTGTGGAGCGCGCGTCT 4922454

		Matches 127; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
QY	321 CTTCCAGGAGATTGCCGCTCCACCATCCCTTACCGCAACCGGAGCTCTCGAGGAAGGT 380	
Db	4922453 TTTTCCTCGAGTCGACGAGACCCCTCATCCGTTCCGGCCGCTCGTCCCGCGGCTCCAGCC 4922394	
QY	381 GGGAGCGTCGCTTCATGCTCCCTCGAGATGTTGCCCTCAACCTGGGGGTCAACCG 440	
Db	4922393 CCGTCCGCGGCTCCGTGGCGCTTCCTCGACGAGGGAAACCGCTGCTGGGGCGGCTCGA 4922334	
QY	441 GCAGACCGCTCCACGCTGGAAGAAGGTCTTTAGAAAAAGGCTGTGGTGGCCACGAGT 500	
Db	4922333 CCAGGCGCTCGACCGCGTCTGATGGCACTGGGTGTGAGTGTGGGCTCGACCTG 4922274	
QY	501 CTTTACCAAAACGTTCAACGGGAGCGCGGCGCATCGGCACCTTTTGGCGGCTCCGCT 560	
Db	4922273 GATGAATGACGCAACGAGTCCGTCGCCCCACGGATCGGACTGCCCCAGGCTGCTGTGGT 4922214	
QY	561 GAGGCGAGGAAAGCAGGCTCACCTTGACGACTACATCTACCCCTGGAGGAACCTCGC 620	
Db	4922213 GGAGTGGCGGAGGCTCCACCACGAGGGCCGCGGCTGCACTGGAAGACCCGCA 4922154	
QY	621 CTTAGACATGGCCACGCGGTCTCTCTTCACTGGGTCAAGGCTTACGAGGACCAAG 680	
Db	4922153 TCTCGTCAAGTGGCGGCGCACCGCCGTTTCTTGGTTCGACGACGAGATCGCGGCCAT 4922094	
QY	681 AATCGCCCCACCTGGAGCTGCTGTCC 709	
Db	4922093 GGACCGCTCTGGTTCGCCCGCAGACCC 4922065	
RESULT 2		
US-10-260-238-721		
; Sequence 721, Application US/10260238		
; Publication No. US20040016025A1		
; GENERAL INFORMATION:		
; APPLICANT: Budworth, Paul R.		
; APPLICANT: Moughamer, Todd R.		
; APPLICANT: Briggs, Steven P.		
; APPLICANT: Cooper, Bret		
; APPLICANT: Glazebrook, Jane		
; APPLICANT: Goff, Stephen A.		
; APPLICANT: Katagiri, Fumiyaki		
; APPLICANT: Kreps, Joel		
; APPLICANT: Provart, Nicholas		
; APPLICANT: Ricke, Darrell		
; APPLICANT: Zhu, Tong		
; TITLE OF INVENTION: PROMOTERS FOR REGULATION OF PLANT EXPRESSION		
; FILE REFERENCE: 60111-NP		
; CURRENT APPLICATION NUMBER: US/10/260,238		
; CURRENT FILING DATE: 2002-09-26		
; PRIOR APPLICATION NUMBER: US 60/325,448		
; PRIOR FILING DATE: 2001-09-26		
; PRIOR APPLICATION NUMBER: US 60/325,277		
; PRIOR FILING DATE: 2001-09-26		
; PRIOR APPLICATION NUMBER: US 60/370,620		
; PRIOR FILING DATE: 2002-04-04		
; NUMBER OF SEQ ID NOS: 6077		
; SEQ ID NO 721		
; LENGTH: 1755		
; TYPE: DNA		
; ORGANISM: Oryza sativa		
; FEATURE:		
; NAME/KEY: N region		
; LOCATION: (917)..(917)		
; OTHER INFORMATION: n = any nucleotide		
; FEATURE:		
; NAME/KEY: N region		
; LOCATION: (1750)..(1750)		
; OTHER INFORMATION: n = any nucleotide		
US-10-260-238-721		
Query Match		5.2%; Score 53.2; DB 15; Length 1755;
Best Local Similarity		50.8%; Pred. No. 1.9e-05;

		Matches 127; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
QY	543 CTTTGGGCGTCCGCTGAGCGAGGAAAGCAGGCTCACCTGGAGGACTACATCTA 602	
Db	246 CGTCTTCGATGGCCGGAGCTCCCGCGAACGACAGCGCATATGGACTACGGGCTCTC 305	
QY	603 CCCTCGAGGAACCTCCCTTAGACATGCGCCCAACGCGGTCTCTCTTCAACTGGGTCAA 662	
Db	306 CCAAGCAAGCTCTTCACTCTGTCATCCGCATCTCCGACCTCCGCTCATCACCGTCCA 365	
QY	663 GGCTTACAGGACCAAGGAAATCCGCCCACTCCCTGAGCTGTGCTTCTTGGGCTCAGG 722	
Db	366 GACCGTCCACGGCAACAAGTTCAGGTTCCGCGTCGAGCCCGCCGACCGCTCGGTAGT 425	
QY	723 GAAAGGTTGATGCCCAACACCAAGACCGTGGCGTTGACCTGGGCTCATCTGGTCT 782	
Db	426 CAAGCAGCAGATCGCCAAAGATCTCCACCCACGACGACACACAGCTCGTCTCTCA 485	
QY	783 CCCCGAGGTG 792	
Db	486 GGGCGAGGTG 495	
RESULT 3		
US-09-758-269-13		
; Sequence 13, Application US/09758269		
; Patent No. US20020104120A1		
; GENERAL INFORMATION:		
; APPLICANT: IUCHI SATOSHI		
; APPLICANT: KOBAYASHI, MASATOMO		
; APPLICANT: SHINOZAKI, KAZUO		
; TITLE OF INVENTION: TRANSGENIC PLANTS CARRYING NEOXANTHIN		
; FILE REFERENCE: 3914-3		
; CURRENT APPLICATION NUMBER: US/09/758,269		
; CURRENT FILING DATE: 2001-01-12		
; PRIOR APPLICATION NUMBER: JP 2001-003476		
; PRIOR FILING DATE: 2001-01-11		
; PRIOR APPLICATION NUMBER: JP 2000-010056		
; PRIOR FILING DATE: 2000-01-13		
; NUMBER OF SEQ ID NOS: 33		
; SOFTWARE: PatentIn Ver. 2.1		
; SEQ ID NO 13		
; LENGTH: 1815		
; TYPE: DNA		
; ORGANISM: Zea mays		
; FEATURE:		
; NAME/KEY: CDS		
; LOCATION: (1)..(1812)		
US-09-758-269-13		
Query Match		5.0%; Score 51.6; DB 9; Length 1815;
Best Local Similarity		45.9%; Pred. No. 5.8e-05;
Matches 177; Conservative 0; Mismatches 209; Indels 0; Gaps 0;		
QY	320 TCCTCCAGGAGATTGCCGCTCCACCATCCCTTACGGCAACCGGAGCTCTGGAGGAAG 379	
Db	155 TCGCCGACCTGCTGCGCTCCAGGAAGCCCGCCCATTCGCCGTCCTCCAGGACACGCG 214	
QY	380 TGGGAGCGTCTCTTTCATGGTCCCTCGAGATGTTGGCCCTCAACTGGGGGTCAACC 439	
Db	215 CGCGCGGAGAAAGCGGAGGGCGGCAAGAGCAGCTCAACTTGTTCAGCGCGCGCGG 274	
QY	440 GCGAGACCGTCCACGCTGGAAGAAGTCTCTTGAGAAAAAGGCGCTGGTGGCCAGCG 499	
Db	275 CGCGCGCTCGACGGCTTCGAGGAAGGTTCTGGCCACGCTCTCGAGCGGCCCGACG 334	
QY	500 TCCTTACCAAAACCGTCAACGGGAGCGCGGCGCATCGGCAACCTTTGGCGGTCCGGC 559	
Db	335 GGCTGCCACGACGGCGGACCGCGCGTGCAGATCGCGGCAACTTCGCGCGCTCGGGG 394	
QY	560 TGAGGCGAGGAAGCGGAGCTCACCTTGACGACTACATCTACCCCTGGAGGAACCTCG 619	

Db 395 AGAGGCGCGCGTGCACGAGTCCCGCTCTCCGGCCGCATCCCGCCCTTCATCGAGGGG 454
Qy 620 CCCTAGACATGGCCAAACGGGTGCTCTCTCACTGGTCAAGGCTTACAGGACCAACG 679
Db 455 TCTACGGCGGCAACGGCGCAACCCCTGCTTCGACCCGTCGCGGGSCACCACTCTTCG 514
Qy 680 GAATCCGCGCCACCTCGAGCTGCTG 705
Db 515 ACGCGAGCGGATGGTGCACGCGCTG 540

RESULT 4
US-10-425-114-35941
; Sequence 35941, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 35941
; LENGTH: 1018
; TYPE: DNA
; ORGANISM: Zea mays subsp. mexicana
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZNROTEOSINTE020D09_FLI
US-10-425-114-35941

Query Match 5.0%; Score 50.8; DB 12; Length 1018;
Best Local Similarity 46.6%; Pred. No. 8.6e-05;
Matches 163; Conservative 0; Mismatches 187; Indels 0; Gaps 0;

Qy 284 AAGCGCTCCAGGACGGGCTCAGAGCTTCTGGAACTCTTGGAACTCTTCCAGGAGATTGCCGCTCCA 343
Db 389 AGGCGCGCAAGCAAAAGCGCGGAGACGACGGAGGCGCGCAAGCAAGCGCGGCGAGA 448
Qy 344 CCATCCCTCAGCAACCGGAGCTCTCGAGAAAGTGGGACGCTGCTTCTCATGTGCC 403
Db 449 CGACGGAGCGCGCAAGCAAGAGAGCGCGGAGACGACGAGGCGCGCAAGCAAGAGCAG 508
Qy 404 CCCTGGAGATTTGGCCCTCAACCTGGGGTACCCGGGACACCGTCCAGCCCTGGAGA 463
Db 509 CCGACGCGCATGGAGCGCGCCCAAGCAGAGAGCGCGCGAGCGCGGACGTACGCCAAGGACA 568
Qy 464 AGGTCCTTGAGAAAAGGCGCTGTGGCCACCGACGCTCTTACCAAAACCGTCAACGGGG 523
Db 569 CGCGGCTCTCGGCAAGAGCAAGTTCGCGGCGCTCATCCAGCAGGCGCACTGAGCAGTGA 628
Qy 524 AGCGCGCGGCATCGGACCCCTTTGGGCGCTCCCGGCTGAGCCAGGAAAGCCAGGCTCA 583
Db 629 AGAGCGCGCGCGCGCGCCCAAGACGCGGTGATGAGCACGCTGGGATGGGCGGGGACA 688
Qy 584 CCCTGGAGCACTACATCTACCCCTGGAGAACCTCGCCCTAGACATGCC 633
Db 689 ACAAGCGGGGACGCCCAACCAACCACTAACTAACCAACCAACCAAGGAC 738

RESULT 5
US-10-156-761-6573
; Sequence 6573, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6573
; LENGTH: 1077
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1077)
US-10-156-761-6573

Query Match 5.0%; Score 50.8; DB 14; Length 1077;
Best Local Similarity 44.3%; Pred. No. 8.7e-05;
Matches 208; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

Qy 323 TCCAGGAGATTGCCCGCTCCACATCCCTACGCAACCGGAGCTCTGGAGAAAGGTGG 382
Db 560 TCGCGGGTGGCTTCGGAGCTCCAGGCGCGCGGATCGCGGCGCGTCAAGATCG 619
Qy 383 GGAAGGTGCTTTCATGTGTCCTTGGAGATTTGGCCCTCAACCTGGGGGTCAACCGGC 442
Db 620 TCGCGGTGACGCTCGCGCGGAGAGATCTCGCCGCGCGGCGCGGCGCGCGACT 679
Qy 443 AGACGCTCCACCGCTCGAGAGAGTCTTGAGAAAAGGCGCTGTGGCCACCGAGCTCC 502
Db 680 ACTCGTCCCTCCGAGAAACACGCGCGCGGAGATCGCGGCTCACCGNACAGAGGGGG 739
Qy 503 TTACCAAAACCGTCAACGGGAGCGCGGCGCATCGGACCCCTTTGGGCGCTCCGCGTGA 562
Db 740 TGGAGCTGCGCGTTCGAGTGCCTGGCGCGCGTGCACCATCGCACCGCTGGGACTCGA 799
Qy 563 GGCAGGGAAGCGAGGCTCACCTGGAGCATATCTACCCCTGGAGAACCTCGCCC 622
Db 800 CGCGCGTGGCGCGCGCACACCGCTCGTGGCATCGGCGGCAAGACACGAGGTCGTGT 859
Qy 623 TAGACATGCCAAACGGCGTCTCTCTTCACTGGGTCAAGGCTCAAGGCTTACAGGACCAAGGAA 682
Db 860 TCAACGCGTGGAGATCTTCACTGGGCGCGCACCTGTGCGGCTCGGTGTAGGCAACT 919
Qy 683 TCGGCGCCACCTCGAGAGTGTCTCTGGTCTCGGCTCAGGGGAAAGGGTGTATGCCCAACA 742
Db 920 CGGACCGCGGAAGACCTCGCGCTCTGGCGGAGCACGTCAGGCGGGGCGCTGGACC 979
Qy 743 CCAAGACCGTGGCGCTTGACCTTGGCTTCATCTCTGCTCTCCCGAGGTG 792
Db 980 TCGCGCGCTGTGAGCGGAGCGGATCGCCCTCGACGGCATCCCGCGGCG 1029

RESULT 6
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262

```

; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      5.0%; Score 50.8; DB 14; Length 9025608;
Best Local Similarity 44.3%; Pred. No. 0.0011;
Matches 208; Conservative 0; Mismatches 262; Indels 0; Gaps 0;

QY 323 TCAGAGAGATTGCCCGCTCACATCCCTACGGCAACCGGAGCTCTGGAGGAGGTGG 382
Db 7895416 TCGCGGGGTGCGCTTCGCGACGCTCCAGCGCCCGCGGATCGCGGCGCGTGAAGATCG 7895475
QY 383 GGACGGTCTCTTCATGTCCTCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCAACCCGCG 442
Db 7895476 TCGCGGTGACGCTCTCGCCGGAGAGGATCGCTCGCCCGCGCGCGGCGCCCGACT 7895535
QY 443 AGACCGTCAACCGCTGGAGAGAGTCTTTGAGAAAAGGCGCTGTGGCCACCGACGTCC 502
Db 7895536 ACCTCGTCCCTCCGAGAACACGCGCCGCGAGATCGCGGGTCAACGACAGAGGGGG 7895595
QY 503 TTCACCAAAACCGTCAACGGGGAGCGCGGGCCATCGGACACCTTTTGGCGCGTCCGGCTGA 562
Db 7895596 TGGACGTGCGCGTCTGAGTCTGTTGGCGCGCGCGCTGACACATCGCACCGCTGGGACTGA 7895655
QY 563 GGCCAGGGAAGCGAGGCTCACCTGGAGAGTACTATCTACCCCTGGAGGAACCTCGCCCG 622
Db 7895656 CGCGCGGTGGCGCGCGCACACGGTCTGCGGATCGGGCGCAAGACACGAGGTCTGTGT 7895715
QY 623 TAGACATGCCAACCGCGTCTCTCTTCACTGGGTCAAGGCTTCAAGGCTTACGAGGACCGGAA 682
Db 7895716 TCACCGCTGTGAGATCTTCCACTGGGCGCGCACCTGTGCGGCTGCGTGTACGGCAACT 7895775
QY 683 TCGCCGCCACCTTGACGCTGTGTCCTCTCTGGGCTCAGGGGAAAAGGTTGATGCCCAACA 742
Db 7895776 CGGACCGCGGAAGGACCTGCCGCTCTGGCGGACGATACGGCGGGCGCGCTGGACC 7895835
QY 743 CCAAGACCGTGGCGTGTGACCTGGGCTCATCTGTCTCTCCCGAGGTG 792
Db 7895836 TCGCGCGCTGTGAGCGGAGCGGATCGCCCTCGACGGGATCCCGCGCGCG 7895885

RESULT 7
US-10-156-761-6076
; Sequence 6076, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HIRIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBATA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697

```

```

; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6076
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(675)
US-10-156-761-6076

Query Match      4.9%; Score 50; DB 14; Length 675;
Best Local Similarity 44.7%; Pred. No. 0.00013;
Matches 237; Conservative 0; Mismatches 290; Indels 3; Gaps 1;

QY 388 GTGCTCTTCATGTCCTCCCTGGAGATGTTGGCCCTCAACCTGGGGGTCAACCGGAGACC 447
Db 106 GACCTCGCCGAGTTCCTTCAGCGCGCCCTGGCGGGGTGACATCTGTGAGCTGCGC 165
QY 448 GTCCACGCTTGAAGAGGTTCCTTTGAGAAAAGGCGCTGTGGCCACCGACGCTCTTCAC 507
Db 166 GACAAAGGATGAGCGCGCGGAGAGCTGGAGCACCTCCAGGTCTTTCGGGAC---GCC 222
QY 508 CAAACCGTCAACGGGAGCGCGGGCCATCGGACACCTTTGGGCGCTTCGGCTGAGGCCA 567
Db 223 TGTGCGCGCCACGCGCAAGCTCTCTCGCGGTCAACGACCGGGCGGACGTCGCGCACGCGGTC 282
QY 568 GGGAAAGCAGGCTCACCTCTGGAGACTATCTTACCTCTGGAGGAACCTCGCCCTTAGAC 627
Db 283 GACTCCGACGTGTCACCTGGGGAGGGGAGCTCCCGTCCCGGGCGCCCGGGCGATC 342
QY 628 ATGGCCAAAGCGGTCTCTCTTCAACTGGGTCAAGGCTTACAGGACCAAGCAATCCGC 687
Db 343 CTGCGCGCCGACGTCCTCTCATCGGCGCTCCACGATGCGGAGCGCGAGCGCGCGCC 402
QY 688 CCACACCTGACGCTGTGTCCTCTGGGCTCAGGGGAAAAGGATGATGCCCAACCAAG 747
Db 403 GCGCTCAGAGGGCGTGGACTTCTTGTACGGGTCTCTGTGGCCACCCGCCCAAG 462
QY 748 ACGTGGCGCTTGACCTGGGCTCATCTGTGTCCTCCCGAGGTGGAGGTTCCAAATC 807
Db 463 CCGCGCGGCAACGCGCGGCTCGACCTGGTTCGGCCACACGCGCGCCCTGGGCAACCGAC 522
QY 808 CCGGCTTATACCTCTCATCTGTAGTACGATTCGCGATCTCTAGATGACGCTGTTCA 867
Db 523 CGCCCTTGTTCGCTATCGGCGCATCGACCTCGGCAATCTCGACGAGTTCTCGAAGCG 582
QY 868 AGACGTTTCTATCAGGCTTGTCTGTGGGCTGTGGCGGAGTCAACCTGCC 917
Db 583 GCGCGCGCGCTGCTGCTGTCGTGCGCGGATCACCAGGCGGACGACCC 632

RESULT 8
US-10-156-761-5431
; Sequence 5431, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: HIRIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBATA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109

```



```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/257,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/289,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 26289
; LENGTH: 1674
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
US-10-282-122A-26289

Query Match          4.8%; Score 49.6; DB 12; Length 1674;
Best Local Similarity 45.7%; Pred. No. 0.00023;
Matches 212; Conservative 0; Mismatches 249; Indels 3; Gaps 1

QY 391 GTCTTCATGTCCTGGAGATGTTGGCCCTCAACTGGGGGTCAACCGGAGACCGTC 450
Db 184 GCCTTCTTGGCCACCGGGCCACCGTGGGCATCTCTCAACAGGAACCGCGCTGAACGAG 243
QY 451 CAGCGCTGGAGAAGGTCTTTCAGAAAAGGGCTGGTGGCCACCGACGCTCTCTTCAACCA 510
Db 244 GACAAAGACCGTTCCGGCAATGTGGAAGAGGGCATGGGGACATCAAGATCAAGCTCGAC 303
QY 511 ACCGTCAACGGGAGCGCGGCCCATCGGACACCTTTGGGCCCGTCCGGCTGAGGCCAGG 570
Db 304 CGCTTCACGAGGTCCCGCAATTGATGCCACCGACTACACCGAGAGCTGATGGAAGAG 363
QY 571 AAAGCCAGGCTCAACCTGGAGCATACATCTACCCCTGGAGGAACCTTCGCCCTAGACATG 630
Db 364 ATGGGTGGCTGCAAGAGGAATCTGAGCACCGCGACGCGTGGGACCTCGACGCGCAGCTC 423
QY 631 GCCAACGGCGTCTCTCTTCAACTGGGTCAAGGCTTACCAGCACCAAGGAATCCGCCCC 690
Db 424 GAGCAGGCCATGGATGCGCTCGGTGTTCGGCGCGCGAC---GAGCGGTAAACCACTTA 480
QY 691 ACCCTGGACGTGTGTCCTCTTGGGCTCAGGGGAAAAGGTGTAGTCCCAACACCAAGACC 750
Db 481 TCGGTGGCAGGCGTCGCGGGTGGCGCTGTGCAACTGCTGTGTTCMAACCCGACCTG 540
QY 751 GTGGCGGTGACCTGGGCTCATCTGGTCTCTCCCGAGGTGGAGCGTTCCAACTCCCG 810
Db 541 TTGTGCTCGACGAGCGGACCAACCACTAGACGCGAAAAGTGTGAGTGGCTCGAACAG 600
QY 811 GCCCTTATCACCCCTCATTTGCTACGTACATTCGCCGATCTCTCTAGA 854
Db 601 CATCTGGCCAGCTACCCCGGTGCGATCTCTGGCGTCAACCA 644

RESULT 10
US-10-282-122A-28540
; Sequence 28540, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; CURRENT FILING DATE: 2003-02-20

```

; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 28540
; LENGTH: 1677
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-10-282-122A-28540

Query Match 4.8%; Score 49.6; DB 12; Length 1677;
Best Local Similarity 45.7%; Pred. No. 0.00023;
Matches 212; Conservative 0; Mismatches 249; Indels 3; Gaps 1;
QY 391 GTCTTCATGTCCTCCCTGGAGATGTTGGCCCTCAACCTGGGGGTACCCGGGAGACCGTC 450
Db 184 GCCTTCTGGCCACCGCGCCACCGTGGGCATCTCTGCAACAGAGAAACCGCGCTGAACGAG 243
QY 451 CAGCGCTGGAAGAAGTCTCTGAGAAAAGGCGCTGGTGCCACCGAGCTCTTCACCAA 510
Db 244 GACAAACCGTTCCGCGGCAATGTGGNAGAGGCAATGGGGACATCAAGATCAAGTCCAC 303
QY 511 ACCGTCAACGGGAGCGCGCGCCATTCGCAACCTTTGGGCGCTCCGGCTAGGCCAGG 570
Db 304 CGTTCAACGAGGTCCGCGCAATGTATGCGCCGACTACACCGAGAGTGATGGAAGAG 363
QY 571 AAAGCCAGGCTCACCTGGACGACTACATCTACCCCTGGAGGAACCTCGCCTAGACATG 630
Db 364 ATGGTTCGCTGCAAGAGGAATGGACACCGCGAGCTGGGACCTCGACGCGAGCTC 423
QY 631 GCCAACGGGTGCTCTCTCTTCAACTGGGTCAAGGCTTACAGGACCAAGGAATCCGCCCC 690
Db 424 GAGCAGGCAATGATGCGCTGCGCTGTCCGCGCGGAC--GAGCGGTAAACCAACCTA 480
QY 691 ACCCTGGACGTGCTGTCTCTGGCTCAGGGGAAAGGATGATCCCAACACCAAGACC 750
Db 481 TCCGTTGGCGAGCTGCGCGGCTGGCTGTGCAAACTGCTGTGTCAAAACCGGACCTG 540
QY 751 GTGGCGGTGACCTGGGCTCTATCTCTGCTCTCCCGAGGTGGAGCGTTCCAAACTCCCG 810
Db 541 TTGTTGCTCGACGAGCGGCAACCAACCTAGACGCGGAAAGTGTGCACTGGCTCGAACAG 600
QY 811 GCCTTATCAACCTCATCTGCTAGTACATTCGCGATCTCTCTAGA 854
Db 601 CATCTGGCCAGTACCCCGGTGCGATCTCTGGCGGTCAACCCAGA 644

RESULT 11
US-10-425-114-32815
; Sequence 32815, Application US/10425114
; Publication No. US20040034868A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 32815
; LENGTH: 898
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMOL7024H02_FLI
US-10-425-114-32815

Query Match 4.8%; Score 49.4; DB 12; Length 898;
Best Local Similarity 44.0%; Pred. No. 0.00022;
Matches 209; Conservative 0; Mismatches 266; Indels 0; Gaps 0;
QY 330 GATTGCCCGCTCCACCATCCCTACGCGCAACCGGAGCTCTGGAGGAAGGTGGGGACGGT 389
Db 1 GATCGAGGACTGCTACATCGTCTCCGGCGAGGACTGCTGGCGCGTGAAGAGCGGTGGGA 60
QY 390 GTCTTCATGTCCTCCCTGGAGATGTTGGCCCTCAACCTGGGGGTACCCGGGAGACCGT 449
Db 61 CGAGTACGGCATCAGGTTCAACATGCGGAGCCAGCATGTCATCCGAGGCTGACCTG 120
QY 450 CCAGCGCTGGAAGAAGTCTCTGAGAAAAGGCGCTGGTGCCACCGAGCTCTTCACCA 509
Db 121 CGTCTCCCCACAGCGCCATGATCGCGCTGGCGAGGAGATGTCGGCGGCAATCCGCA 180
QY 510 AACCGTCAACGGGAGCGCGCGCCATTCGGCAACCTTTGGGCGCTCCGGCTAGGCCAGG 569
Db 181 CGTGGCGCGGAGGAGAGCGTCCGCTCAACCGGAGTCCGCGTCAAGGTCAAGTCCGG 240
QY 570 GAAAGCCAGGCTCACCTGGAGGACTACATCTACCCCTGGAGGAACCTCGCCCTAGACAT 629
Db 241 TGGCGGAGGGCGGCTTGTGTAAGACATCTCTGTCGGGCGCTCAGCTCCACACAT 300
QY 630 GGCCAAACGGCGTCTCTCTTCAACTGGGTCAAGGCTTACAGGACCAAGGAATCCGCCC 689
Db 301 GAAGTGGCTTCTGATGATACCGGCAACTACGGGAGCACCAGCAACACGCTCCGACCC 360
QY 690 CACCTGGAGCTGCTGCTCTCTGGGCTCAGGGGAAAGGATGATGCCCAACACCAAGAC 749
Db 361 CAACGCGCATCCCGAGGTACCGGCAATCACTACGCGAGGTTCGCGGAGAACGTGAC 420
QY 750 CGTGGCGGTGACCTGGGCTCATCTCTGCTCTCCCGAGGTGGAGCGTTCCAAA 804
Db 421 CATGGCGGAGGATGAGGGGATCCCTAAGGACCCCTACACCGGATCTGCATA 475

RESULT 12
US-10-369-493-42225
; Sequence 42225, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 42225
; LENGTH: 1803
; TYPE: DNA
; ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-42225

Query Match 4.8%; Score 49.2; DB 15; Length 1803;
Best Local Similarity 48.3%; Pred. No. 0.0003;
Matches 138; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

Qy 426 CTTGGGGTCCACCGGAGACCGTCCACGCTGGAAGAGTCTCTGAGAAAAGGCGCT 485
Db 198 CTTCTCTGTCACCGACTTATGACCGCGGCGGACCTATACGGCGAAGAACCGCGT 257
Qy 486 GGTGGCCACCGACGCTCTTCCAAACCGTCAACGGGAGCGCGGGCCATCGGACCCCT 545
Db 258 CGTGGCCACCGACCGCGGGGTACAGTACAGGAGCTCGCGACCGCGCGCGCGCTT 317
Qy 546 TTGGGCGGTGCGGTGAGCGGAGGAGGAGCGGCTCACCTGAGAGTACATCTACCC 605
Db 318 TTCGGCGGTCTCCAGGCGGCGGCGATCGACAGGGGAGCGCGTCCCGTCTCGACCC 377
Qy 606 CTGAGGAACCTCGCCCTAGACATGGCCAAACGGCGGTCTCTCTTCAACTGGGTCAAGGC 665
Db 378 GAACACCCACTACACCTCGAAGCGGCTACGGCGGATGAGCTGGGGGCGCATCCACAC 437
Qy 666 CTACAGGACCAAGGAATCCGCCCAACCGTGGAGTGTGTCCTC 711
Db 438 GCCACTGAATACCGGTCTACGCGCGGAGCTACGCGGTATCATCTC 483

RESULT 13
US-10-425-114-26418
; Sequence 26418, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 26418
; LENGTH: 1045
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4172-028-B4_FLI
US-10-425-114-26418

Query Match 4.8%; Score 48.8; DB 12; Length 1045;
Best Local Similarity 45.7%; Pred. No. 0.00034;
Matches 170; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

Qy 334 GCCCGCTCCACCATCCCTACGGCAACCGGAGCTCTGAGGAAGGTGGGACCGTCTC 393
Db 384 GGGCGGTCCAGGAAGCGCGCCCATTCGCTCCAGGGCACTCCGCGGCGCGGAGAAA 443
Qy 394 TTTATGTTCCCTCGAGATGTGGCCCTCAACCTTGGGGGTCAACCGGCGAGACCGTCCAC 453
Db 444 GGGAGCGCGGCAAGNAGAGTCACTGTTTCAGAGCGCGCGCGCGCGCGCTCGAC 503
Qy 454 GCGTGAAGAAGGTCTTTCAGAAAAGGGCTGGTGGCAACCGAGTCTCTTCAACCAACC 513
Db 504 GCGTTGAGGAAGGTTCGTGGCCAAAGTCTCTGGAGCGCGCCCAACGGGTGCCCCAGCAG 563
Qy 514 GTCAACGGGAGGCGCGGCGCATCGGCACCTTTGGGCGGTCCGCTGAGGCGGAGAAA 573

Db 564 GCCGACCGCGCGGTGCAGATCGCGGCACTTCGCGCCGTCGGGAGAGCGCGCGCTG 623
Qy 574 GCCAGGCTCACCTGGACGACTATATCTACCCCTGGAGGAACCTCGCCCTAGACATGCC 633
Db 624 CACGAGTCCCGGTCTCCGCGCGCATCCCGGCGCTTATCGACGCGGTCTACGCGCGAAC 683
Qy 634 AACGGGTGCTCTCTTCAACTGGGTCAAGGCTACAGGACCAAGGAATCGCCCCACC 693
Db 684 GCGCGCAACCCCTGTTTCGACCCCGTCCCGGGGACCACTCTTCGACGGGACGGCATG 743
Qy 694 CTGGACGTGCTG 705
Db 744 GTGACGCGCTG 755

RESULT 14
US-10-369-493-34210
; Sequence 34210, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 34210
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Shingomonas aromaticivorans
US-10-369-493-34210

Query Match 4.7%; Score 48.6; DB 15; Length 960;
Best Local Similarity 45.0%; Pred. No. 0.00038;
Matches 183; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

Qy 396 CATGTCCCTCGGAGATGTGGCCCTCAACTGGGGGTGTCACCGGACAGCGTCCACGC 455
Db 204 CATCGCAGCGCGGACGTGATCATCTCACCGCGCGCTCGCCGCAAGCCCGGATGAG 263
Qy 456 CTGGAAGAAGTCTTTGAGAAAAGGCGCTGTGGCCACCGAGCTCTTCAACCAACCGT 515
Db 264 CCGGACGACCTGCTCGGATCAACCTCAAGGTGATGAAGGCGCTCGGGAAGGATCCG 323
Qy 516 CAACGGGAGCGCGGCGCATCGGCACCCCTTTGGGCGCTCGGCTGAGGCCAGGAAAGC 575
Db 324 CGACAAGCGCGCGACGCGTTCGTGATCTGCATCAACCCCGCTCGAAGCGATGCTG 383
Qy 576 CAGGTCACTCGGACGACTATATCTACCCCTGGAGGAACCTCGCCCTAGACATGCCAA 635
Db 384 GGGCTTCGCGAGTTCGCGGCGCTCGCCGCGCAACAGGTGCTCGGATGCGCGCGTGT 443
Qy 636 CCGGTGCTCTCTCTTCAACTGGGTCAAGGCTACAGGACCAAGGAATCGCCCCACCCT 695
Db 444 CGACTCGCGCGGTTCAGCACGTTCTGGATGGGAATTCGGGCTCTCGATCCGCGACGT 503
Qy 696 GGAAGTCTGCTCTCTGCGGTGAGGGGAAAGGGGTGATGCCCAACCAAGACCGTGGC 755
Db 504 GAACACGTTCTGTTCTCGCGGCGCACGCGGACCATGTTTCCGGTCAACCCAGTACTGAC 563
Qy 756 GGTGACCTGGGCTCATCTGCTCTCCCGAGGTGGAGCGTTCCA 802
Db 564 CGTCAACGGCATCCGGTCCCGACCTCTGTCAGATGGGCGCTGTCCA 610

RESULT 15
US-10-200-562-190
; Sequence 190, Application US/10200562
; Publication No. US20030165819A1
; GENERAL INFORMATION:
; APPLICANT: McGowan, Patrick
; APPLICANT: Hosken, Nancy A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF HERPES SIMPLEX VIRUS INFECTION
; FILE REFERENCE: 210121.538C2
; CURRENT APPLICATION NUMBER: US/10/200,562
; CURRENT FILING DATE: 2002-07-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 190
; LENGTH: 9369
; TYPE: DNA
; ORGANISM: HSV2
US-10-200-562-190

	Query Match	4.7%	Score 48.6;	DB 14;	Length 9369;
	Best Local Similarity	43.9%	Pred. No. 0.00073;		
	Matches 207;	Conservative 0;	Mismatches 264;	Indels 0;	Gaps 0;
Qy	355	GGCAACCGGAGCTCTGGAGGAAGTGGGGACGGTCTTCATGTTCCCTCGGAGATG	414		
Db	3643	GCATCGACAGGCTACAGACCTTGGGTCGGCGCGCGACCTGGTAGTCCAGGCGCGC	3702		
Qy	415	TTGGCCCTCAACCTGGGGTCAACCGGCAGACCGTCCAGCTGGGAAGTCTTTGAG	474		
Db	3703	CGGGCTTGCCCGACGCGAGGGGACCATGCGCCCTCATGATGCCGCGCCGCGCG	3762		
Qy	475	AAAAAGGGCTGTGTGCCACCGACGTCCTTCAACCAACCGTCAACGGGAGCGCCGGGCC	534		
Db	3763	ACGACCGCTGCTCGGGAAGCCTCGCGGGGCACAGGCTGGGTTCGGGGCCCTGCTGCAC	3822		
Qy	535	ATCGGCACCTTTGGGCGCTCGGCTGAGGCCAGGAAGCCAGGCTCACCTGGACGAC	594		
Db	3823	GCCGAAGGACGGCCGGGACCATTTCCCGAGGGCGCGCCCTGAGGAGCTGGGCAAG	3882		
Qy	595	TACATCTACCCCTGGAGGAACCTCGCCCTAGACATGGCCACGGGTGCTCTCTTCAAC	654		
Db	3883	GTCTCGGGCCACCGCGACCGCGCGCGACGAACTTGGCGCGCGTCCCGACCTGAGG	3942		
Qy	655	TGGGTCAAGGCTTACGAGGACCAAGGATCCGGCCCGACCTGGAGCTGTGTCTCTGG	714		
Db	3943	GCGAAGATGGCGGCCCGAGCGCGCCGGGCGACGAGCGCTGGGCGCGCGCGCTCGAG	4002		
Qy	715	GCTCAGGGGAAGGCTGATGCCAACACCAAGACCGTGGCGGTTGACCTGGGCTCATC	774		
Db	4003	GCGGCGCTGACCGCGCTCGAGAACCGCGCGGAGTTGACGTGTGAGTGGCGCGCTG	4062		
Qy	775	CTGGTCTCCCGAGGTGGAGGTTCCAAACTCCCGGCCCTTATCACCTC	825		
Db	4063	CAGGCGCTGGCGGGCACGACGCGTACAAACCCCGGGACTTCCGCAAGCGC	4113		

Search completed: March 18, 2004, 10:21:33
Job time : 259.943 secs